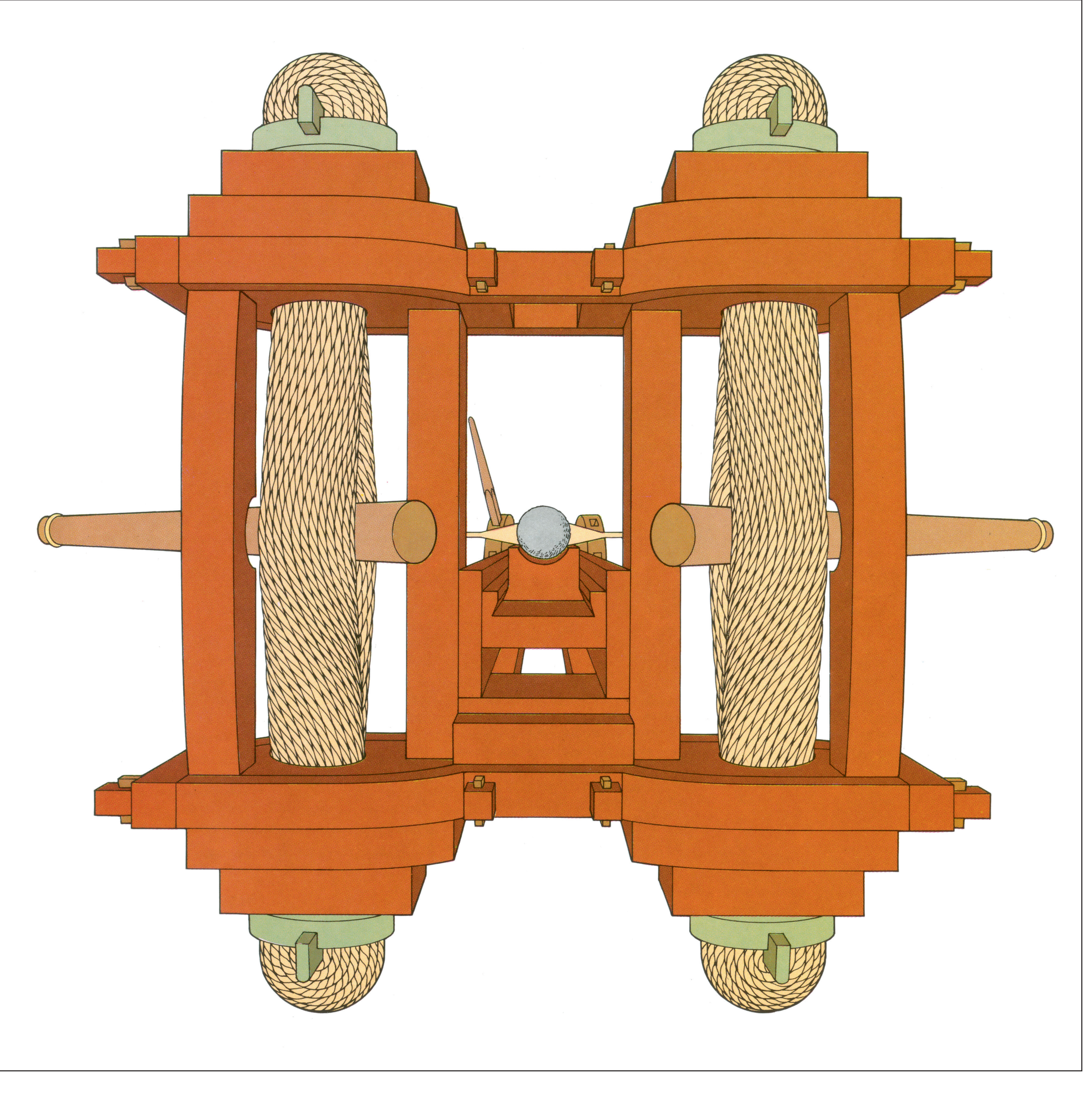
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



CATAPULTAS ANTIGUAS

Mayo 1979

200 PTAS.

Copyright © 1979 Prensa Científica S.A.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 32
Mayo 1979

INVESTIGACION Y CIENCIA
Edición en español de SCIENTIFIC AMERICAN

SCIENTIFIC AMERICAN

COMITE DE REDACCION Gerard Piel (Presidente), Dennis Flanagan, Francis Bello, Philip Morrison, Judith Friedman, Brian P. Hayes, Paul W. Hoffman, Jonathan B. Piel, John Purcell, James T. Rogers, Armand Schwab, Jr., Jonathan B. Tucker y Joseph Wisnovsky

DIRECCION EDITORIAL Dennis Flanagan
DIRECCION ARTISTICA Samuel L. Howard
PRODUCCION Richard Sasso
DIRECTOR GENERAL Donald H. Miller, Jr.

INVESTIGACION Y CIENCIA

DIRECTOR Francisco Gracia Guillén
REDACCION José María Valderas Gallardo (Redactor Jefe)
Encarna de Blas (Secretaria de Redacción)
PRODUCCION César Redondo Zayas
PROMOCION Y PUBLICIDAD Elena Sánchez-Fabrés
PROMOCION EXTERIOR Pedro Clotas Cierco

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

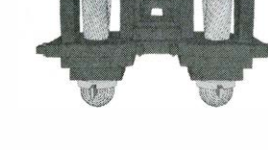
Colaboradores de este número:

Asesoramiento y traducción:

Manuel Puigcerver: *Objetos Apolo;* Nicolás Herraiz: *Metástasis cancerosas;* Pedro Pascual: *El espectro del hidrógeno atómico;* Justiniano Casas: *La percepción de los blancos y negros de las superficies;* Cristóbal Mezquita: *Canales iónicos en la membrana de la célula nerviosa;* Joandomènec Ros: *Delfines;* Julio Samsó: *Catapultas antiguas;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio;* Carlos Oppenheimer: *Ciencia y sociedad.*

Libros:

Ramón Margalef, Jorge Tamarit, Luis Vila López y Andrés González Navarro



LA PORTADA

El dibujo de la portada ilustra la visión frontal de una catapulta lanzapiedras construida según las especificaciones dejadas por el ingeniero militar romano Vitrubio, hacia el año 25 a. C. Este modelo particular de arma se basaba en el principio de torsión. Utilizaba ovillos muy tirantes de fibras elásticas, sacadas de tendones animales, crines de caballo o pelo humano; las fibras recibían un refuerzo suplementario, cuando el arma estaba plenamente tensada, mediante un brazo rígido de arco (véase "Catapultas antiguas", por Werner Soedel y Vernard Foley, en este mismo número). La pieza básica era el mango (que aparece en la parte inferior central de la estructura en *H* de la portada), una barra compuesta que constituía el eje principal del arma. En su parte superior había una ranura sobre la que se ensamblaba una segunda barra, la corredera, que podía moverse hacia adelante y hacia atrás. Esta corredera llevaba, sobre su superficie trasera, un mecanismo a base de retén y gatillo, estructura que permitía sujetar y liberar la cuerda del retén.

Suscripciones:
Prensa Científica, S.A.
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2200 pesetas
Extranjero:
Un año (12 números): 36 U.S.$
Ejemplar atrasado ordinario:
225 pesetas
Ejemplar atrasado extraordinario:
340 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49; Barcelona-10

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239; Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovín
Avda. de Moratalaz, 137; Madrid-30
Tel. 439 41 50 ext. 468
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 3.ª; Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Ilustración de la portada de Dan Todd

Página	Fuente
9	Richard F. McCrosky y Cheng-Yuan Shao, Harvard College Observatory (*arriba*); Eleanor F. Helin, Instituto de Tecnología de California (*abajo*)
10	National Aeronautics and Space Administration
11	Andrew Tomko
12-13	Dan Todd
14	Albert Miller
15	Departamento de Energía, Minas y Recursos, Ottawa
16-20	Albert Miller
23	Garth L. Nicolson, Universidad de California en Irvine
24-32	Carol Donner
34-35	Gerhard Herzberg (*arriba*), Jon Brenneis (*abajo*)
36-47	Gabor Kiss
59	Alan L. Gilchrist, Universidad estatal de Nueva York en Stony Brook
60	Alan L. Gilchrist, Universidad de Nueva York en Stony Brook (*arriba*); Ilil Arbel (*abajo*)
61-68	Ilil Arbel
73	Alexander von Muralt, Universidad de Berna
74-79	Jerome Kuhl
83	Photo Researchers, Inc. (*arriba*), Rapho-Photo Researchers Inc. (*abajo*)
84-86	Tom Prentiss
87	Bernd Würsig, Universidad de California en Santa Cruz
88-90	Tom Prentiss
91	Photo Researchers, Inc. (*arriba izquierda y abajo*) Bernd Würsig, Universidad de California en Santa Cruz (*arriba, derecha*)
93-100	Dan Todd
102-108	Antonio Martínez
110-113	Ilil Arbel
116-121	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona
Printed in Spain - Impreso en España

Copyright © 1979 Scientific American, Inc., 415 Madison Av., New York, N.Y. 10017.

Copyright © 1979 Prensa Científica, S.A., Calabria, 235-239 - Barcelona-29 (España).

Reservados todos los derechos. Prohibida la reproducción en todo o en parte por ningún medio mecánico, fotográfico o electrónico, así como cualquier clase de copia, reproducción, registro o transmisión para uso público o privado, sin la previa autorización escrita del editor de la revista

El nombre y la marca comerciales SCIENTIFIC AMERICAN, así como el logotipo distintivo correspondiente, son propiedad exclusiva de Scientific American, Inc., con cuya licencia se utilizan aquí.

Los autores

GEORGE W. WETHERILL ("Objetos Apolo") es director del Departamento de Mangnetismo Terrestre de la Institución Carnegie de Washington. Estudió en la Universidad de Chicago, en donde obtuvo el grado de Doctor en Física en 1953. Después entró a formar parte del personal técnico del Departamento de Magnetismo Terrestre. En 1960 fue nombrado profesor de Geofísica y Geología en la Universidad de California en Los Angeles, en la que desempeñó desde 1968 hasta 1972 el cargo de director del Departamento de Ciencia Espacial y Planetaria. Regresó al Departamento de Magnetismo Terrestre en 1975.

GARTH L. NICOLSON ("Metástasis cancerosas") enseña biología celular y es director adjunto del programa de oncología de la Universidad de California en Irvine. Tras graduarse en Los Angeles, se licenció en biofísica en la Universidad de Hawai, doctorándose en biología celular por la Universidad de California en San Diego. Trabajó en el laboratorio de S. J. Singer y contribuyó a la formulación del modelo en mosaico fluido propuesto para la membrana biológica. Su investigación actual se centra en el estudio de la superficie celular de las células normales y tumorales.

THEODOR W. HÄNSCH, ARTHUR L. SCHAWLOW y GEORGE W. SERIES ("El espectro del hidrógeno atómico") han usado láser para estudios ópticos de alta resolución de átomos simples. Hänsch es profesor de física en la Universidad de Stanford. Nacido en Alemania Occidental, estudió en la Universidad de Heidelberg, donde obtuvo brillantemente su doctorado en física en 1969. Desde 1973 a 1975 Hänsh fue becario de la fundación Alfred P. Sloan Research. Shawlow es profesor J. G. Jackson – C. J. Wood de física en Stanford. Se doctoró por la Universidad de Toronto en 1949. Fue coautor con Charles H. Townes del primer trabajo en que se describía un láser. En 1977 se le concedió la tercera beca internacional Marconi, que le fue entregada en Estocolmo por el rey de Suecia. Series enseña física en la Universidad de Reading. Estudió en la Universidad de Oxford, donde obtuvo su Ph. D. en espectroscopía atómica. Después fue lector de física en Oxford y pasó a Reading, en 1969.

ALAN L. GILCHRIST ("La percepción de los blancos y los negros de las superficies") es profesor auxiliar de psicología en la Universidad del estado de Nueva York en Stony Brook. Se formó en el Instituto para Estudios de Cognición en la Universidad de Rutgers, donde se doctoró en 1975. El escribe: "Mi verdadero interés reside en los problemas relacionados con la naturaleza del proceso del pensamiento. Decidí estudiar la percepción, porque creo que representa el campo de investigación más inmediato a la consciencia en el que caben progresos substanciales".

RICHARD D. KEYNES ("Canales iónicos en la membrana de la célula nerviosa") es profesor de fisiología de la Universidad de Cambridge. Nació en Londres en 1919, educándose en el Trinity College de Cambridge. Durante la Segunda Guerra Mundial trabajó en radar y sonar para la Royal Navy como científico civil, regresando a Cambridge en octubre de 1945 donde se graduó en Fisiología en el mes de junio. Cuatro años más tarde obtuvo el grado de doctor defendiendo sus tesis sobre la determinación de los movimientos iónicos durante la actividad nerviosa. Hasta 1960 permaneció en Cambridge primero como investigador postdoctoral en el Trinity College y luego como profesor en Peterhouse. Se trasladó después al Instituto de Fisiología Animal del Consejo de Investigación Agrícola de Babraham, donde fue director desde 1965 a 1973.

BERND WÜRSIG ("Delfines") es un becario postgraduado en el Center for Coastal Marine Studies de la Universidasd de California en Santa Cruz. Escribe: "Soy hijo de padres refugiados en Alemania Oriental. Nos establecimos en Ohio en 1956, y aunque no llegué a tocar el agua del mar hasta 1969 (en Livorno, Italia, para ser exactos), mucho antes de esta época se desarrolló en mí un gran interés por la biología marina, y en particular por los cetáceos, grandes y pequeños. Más tarde, en la Ohio State University, abandoné unos inicios prometedores en biología del desarrollo simplemente porque sabía que nunca sería muy feliz en el laboratorio". Würsig realizó sus estudios de grado sobre comportamiento de los cetáceos en la Universidad estatal de Nueva York en Stony Brook, y obtuvo su doctorado en 1978.

WERNER SOEDEL y VERNARD FOLEY ("Catapultas antiguas") son profesores de la Universidad de Purdue. Soedel enseña ingeniería mecánica. Nació en Praga y pertenecía a la minoría alemana de los Sudetes que fue deportada a Alemania en 1945. Tras realizar estudios en un pueblecito de la Baviera sudoccidental, siguió los cursos de ingeniería mecánica en la Escuela estatal de Ingenieros de Francfurt donde se graduó en 1957. En 1963 emigró a los Estados Unidos y cursó estudios en Purdue donde se doctoró en 1967 pasando, a continuación, a formar parte de su profesorado. Allí enseña e investiga en el área de la mecánica teórica y experimental. Soedel se ha interesado por la física de las catapultas desde su época de estudiante, cuando tuvo ocasión de ver demostraciones de tiro con algunas de las catapultas reconstruidas en Saalburg, una fortificación romana convertida en museo en las colinas de Taunus, al norte de Francfurt. Foley es profesor adjunto de Historia en Purdue. Obtuvo su título de graduado en el McPherson College, el de licenciado en la Universidad de Kansas y el de doctor en la Universidad de California, en Berkeley, en 1970. Obtuvo los dos últimos títulos dentro de la especialidad de Historia de Europa, particularmente en Historia de la Ciencia y de la Tecnología.

ANTONIO MARTINEZ HERNANDEZ ("Inmunohistoquímica en microscopía electrónica") nació en Calahorra (Logroño) en 1944. Cursó estudios de medicina en la Universidad Complutense (Madrid), donde se graduó en 1968. Hizo su internado en Tucson (Arizona). Realizó su residencia en anatomía patológica en la Facultad de Medicina de la Universidad de Colorado (Denver), de donde fue Assistant y Associate Professor. Desde julio de 1978 es Full Professor en el Departamento de patología de la Facultad de Medicina de Hahnemann (Philadelphia). Sus áreas de interés son: la anatomía patológica cardiovascular y la inmunohistoquímica del tejido conjuntivo.

Hace...

José M.ª López Piñero

... *cuatrocientos años*

Apareció en París la segunda edición del tratado de física de Benito Perera, titulado *De communibus omniun rerum naturalium principiis et affectionibus.* Nacido en Ruzafa, que hoy es un barrio de la ciudad de Valencia, Perera residió la mayor parte de su vida en Roma, tras ingresar muy joven en la Compañía de Jesús. Su tratado de física, publicado por vez primera en 1562, alcanzó después una difusión extraordinaria a través de nueve reediciones en Francia, Alemania, Inglaterra e Italia. Es una excelente exposición, en la que se defiende el valor de la experiencia frente a la autoridad de los clásicos: "En física, tengo en cuenta, en primer lugar, el juicio de los sentidos, investigado y confirmado con larga experiencia y diligente observación; en segundo lugar, la razón; en último, la autoridad de los filósofos". Ello no impide, sin embargo, que el estudio dedicado al movimiento local sea un mero retorno a las doctrinas tradicionales anteriores a las críticas nominalistas que, en España, habían culminado con la formulación de la ley de caída de graves por Domingo de Soto. El libro de Perera, junto a los del cordobés Francisco de Toledo y al de Domingo de Soto, aparece citado en las notas juveniles de Galileo.

Entre las demás obras de Perera, destaca *Adversus fallaces et superstitiosas artes* (1603), uno de los títulos más importantes de la literatura antisupersticiosa de la época. Tuvo ocho ediciones latinas, en Italia, Francia y Alemania, así como dos en lengua inglesa.

... *trescientos años*

Organizó en Barcelona su herbario Jaume Salvador Pedrol, la figura más importante de una notable dinastía de botánicos. Boticario de profesión, inició su formación como naturalista junto a su padre, Joan Salvador Boscà, completándola más tarde en Montpellier. Aunque no llegó a publicar ninguna obra, ocupa un lugar destacado en la decaída botánica española de la época, por la importante colaboración que prestó a Tournefort en sus viajes para estudiar la flora de Cataluña y Valencia (1681) y, más tarde, la de toda la Península Ibérica (1688). Las contribuciones de Salvador fueron recogidas por el botánico francés en sus *Elémens* (1694) y en su *Corollarium institutionum rei herbariae* (1703), donde lo llama "Fénix de su patria". La clasificación de Tournefort, la más avanzada entre las anteriores a Linneo, fue introducida en España por el boticario barcelonés, que la utilizó para ordenar, no solamente su herbario antes citado, sino un jardín botánico que fundó en Sant Joan Despí, junto a Barcelona. Colmeiro consideró que este último era, "si no el primero, el más rico por lo menos y el más propio de su objeto, que hasta entonces se había conocido en España".

... *doscientos años*

Juan González Figueras construyó seis juegos de instrumentos científicos con destino a las minas de Almadén, bajo la dirección del ingeniero militar Miguel Sánchez. González Figueras ha sido considerado por Víctor Navarro como una de las principales figuras de la España del siglo XVIII en el terreno de la instrumentación científica. Carecía de estudios superiores y se formó como tornero junto a su padre, llegando a maestro en este oficio. Construyó instrumentos para numerosas instituciones científicas y técnicas, entre ellas, la Real Academia del Regimiento de Reales Guardias, la Escuela de Náutica de Barcelona y la Universidad de Salamanca. El claustro de esta última, tras encargarle varios aparatos e instrumentos de física experimental, le nombró, en 1775, "maestro maquinista e hidráulico". Especial relación mantuvo con la Real Academia de Ciencias Naturales y Artes de Barcelona, de la que fue nombrado miembro en 1776 y "custode de las máquinas" diez años más tarde. También trabajó para los colegios de Cirugía fundados por el movimiento reformista de la profesion y la enseñanza quirúrgicas encabezado por Virgili y Gimbernat. Además de instrumentos y aparatos de física, máquinas eléctricas y neumáticas, microscopios, etc., construyó modelos del oído, del globo ocular, del corazón y de otros órganos del cuerpo humano. Por su gran prestigio, una real cédula le concedió varias exenciones, así como "la facultad de hacer trabajar en su casa a los mancebos de cualquier gremio y permiso de poner tienda pública para vender sus máquinas en cualquier parte del Reino".

... *cien años*

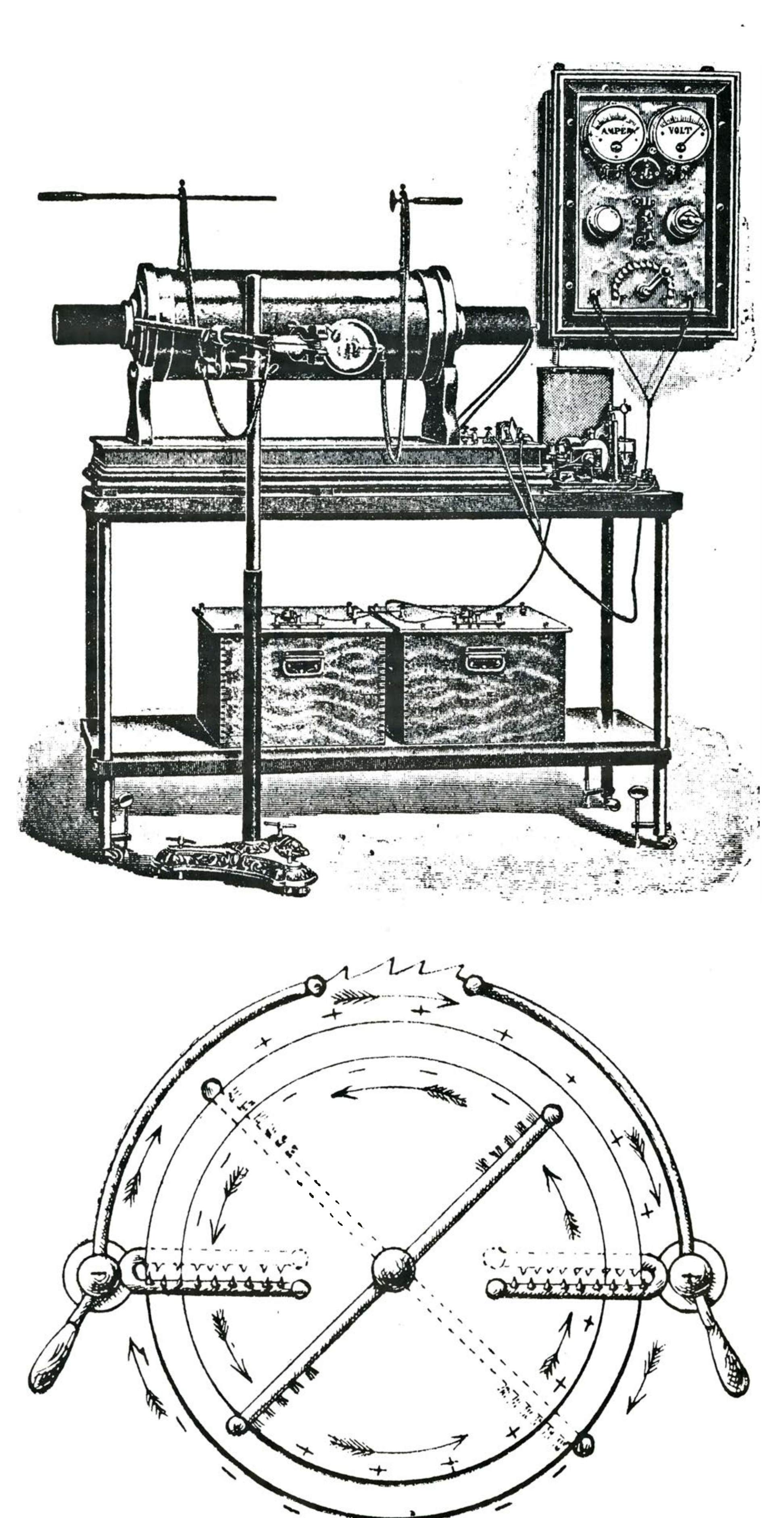
Se doctoró en ciencias fisicoquímicas Vicente Peset Cervera. Miembro de una notable dinastía de médicos valencianos, había obtenido cuatro años antes otro doctorado en medicina. Consagró su vida al cultivo de la química y, en general, de las disciplinas experimentales, principalemente en relación con sus aplicaciones médicas.

Desde la cátedra de terapéutica de la Facultad de Medicina de Valencia, que ocupó a partir de 1892, Peset Cervera contribuyó a la introducción en España de numerosas novedades científicas y técnicas. Su monografía sobre *La fermentación en Fisiología y Patología* (1880) incluye una de las primeras síntesis de la naciente microbiología publicadas en nuestro país. Por otra parte, se ocupó de las aplicaciones de la electricidad a la terapéutica y al diagnóstico, tema sobre el que escribió un excelente manual (1882). También fue uno de los primeros españoles que trabajó con los rayos X. Comenzó a manejarlos pocos meses después de la célebre comunicación de Wilhelm Roentgen y, dos años más tarde, presentó al X Congreso Internacional de Higiene y Demografía un notable trabajo experimental acerca de la detección de adulteraciones medicamentosas con las nuevas radiaciones. Aparte de desarrollar una incansable actividad de analista químico, sobre todo de aguas potables y mineromedicinales, pero también de medicamentos, alimentos e incluso productos industriales, el catedrático valenciano fue autor de una obra escrita extraordinariamente amplia. Publicó, en revistas profesionales y generales, un gran número de artículos científicos y de divulgación sobre materias muy diversas. La historia de la medicina y de la ciencia fue un tema al que prestó atención a lo largo de toda su vida. Gran difusión alcanzaron sus traducciones castellanas, con abundantes adiciones, del gran tratado de análisis químico de Carl Remigius Fresenius (1885-87) y del resumen de bioquímica de Charles Adolphe Wurtz (1891). Mención aparte merece su manual de terapéutica, aparecido por vez primera en 1894 y reeditado en la década siguiente. Fue el de mayor rigurosidad y altura publicado en la España de la época y significó, entre otras cosas, la plena incorporación de la moderna farmacología experimental a nuestro país.

En su *Terapéutica* dejó escrito: "Toda instalación completa de rayos X, fija o portátil, requiere un manantial eléctrico apropiado, aparatos de medida y muchos accesorios, siendo costosas. Por lo respectivo a la energía puede ser de cualquier origen; pero son más recomendables las máquinas electro-estáticas con alta tensión y débil intensidad, sobre todo la de Wirmshurst, porque los demás medios exigen transformadores que sólo aprovechan 20 por ciento de los vatios suministrados, y las pilas primarias con su mínima fuerza electromotora son incapaces de alumbrar una ampolla. La máquina de Wirmshurst contendrá 6 o 12 discos de ebonita de 50 cm diámetro, sin conductores ni sectores metálicos encolados que cambian el sentido de la corriente, pero con pinceles frotadores y una grande velocidad de rotación a merced de un motor, conservándola en urna de vidrio abierta durante el trabajo para que el ozono producido no la destruya; su polo negativo, que da luces violáceas en el colector correspondiente, se une al cátodo de la ampolla y el otro al ánodo".

... cincuenta años

Eduard Fontseré Riba fundó un observatorio meteorológico en la cumbre del Montseny y organizó en Barcelona una reunión de la Comisión Internacional para el Estudio de las Nubes. Formado en la Universidad de Barcelona, donde estudió la licenciatura de ciencias fisicomatemáticas, Fontseré cultivó la astronomía y la sismología, publicó numerosos textos destinados a la docencia y a la divulgación y, sobre todo, fue una sobresaliente figura de la meteorología. Inició sus estudios sobre esta última disciplina en 1896, organizando la red pluviométrica de Cataluña y Baleares, así como la publicación regular de observaciones meteorológicas en el *Boletín* de la Granja Agrícola Experimental de Barcelona. En 1912, fue nombrado director de la sección meteorológica del Observatorio Fabra y, al año siguiente, fundó la Estación Aerológica de Barcelona, base del Servicio Meteorológico de Cataluña, que sería su principal contribución. Su labor consiguió reconocimiento internacional a partir de la década de los veinte, durante la cual fue nombrado miembro de la Comisión Permanente de la Red Meteorológica Mundial y de otros organismos internacionales, entre ellos el antes citado. De sus abundantes publicaciones, anotaremos un *Atlas elemental de núvols* (1925) y unos *Elementos de meteorología* (1943).



Instalación de rayos X y esquema de la máquina de Wirmshurst. Figuras de la segunda edición (1905) de la **Terapéutica** *de Vicente Peset Cervera*

Objetos Apolo

Son cuerpos asteroideos cuyas órbitas cortan a la órbita de la Tierra. Se conocen actualmente unos 30 de ellos. Pueden ser núcleos de cometas que han perdido sus componentes volátiles

George W. Wetherill

En 1937, un cuerpo de alrededor de un kilómetro de diámetro, al que después se llamó Hermes, pasó a menos de 800.000 kilómetros de la Tierra, distancia no mayor del doble de la distancia a la Luna. No se le ha vuelto a ver. Aproximadamente una vez cada siglo es de esperar que un objeto similar pase a una distancia de la Tierra menor que la distancia lunar. Y una vez cada 250.000 años, por término medio, un tal cuerpo y la Tierra chocarán. El impacto del choque liberará una energía equivalente a 10.000 bombas de hidrógeno de 10 megatones y producirá un cráter de unos 20 kilómetros de diámetro. Afortunadamente, tales catástrofes son tan infrecuentes que no se ha registrado ninguna de ellas en la historia de la humanidad.

Hermes, el asteroide que se descubrió cruzando la órbita terrestre en 1937, sólo era el tercer objeto de su clase que se identificaba. El primero de ellos había sido descubierto cinco años antes, en 1932, por Karl Reinmuth, de la Universidad de Heidelberg, en el curso de una búsqueda fotográfica de asteroides ordinarios. Sin duda que anteriormente se habían fotografiado objetos similares, pero nunca habían sido correctamente identificados. El rápido movimiento de un cuerpo que se acerca a la Tierra deja en la placa fotográfica una traza que no es fácil de distinguir de la traza de un meteoro o de un defecto fotográfico. Reinmuth fue el primero que demostró que el objeto que había fotografiado era un asteroide cuya trayectoria elíptica cortaba a la órbita terrestre antes de que alcanzase su perihelio (punto más próximo al Sol). Se dio al objeto la designación provisional 1932 HA, y después se le llamó Apolo, como la divinidad mitológica griega. El término "objetos Apolo" se aplicó más tarde a la totalidad de los cuerpos cuyas órbitas cortan a la terrestre.

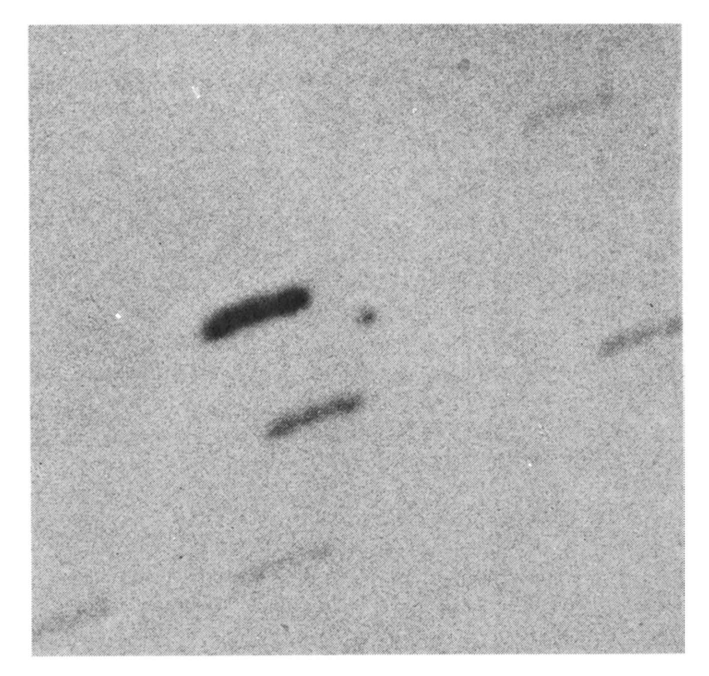
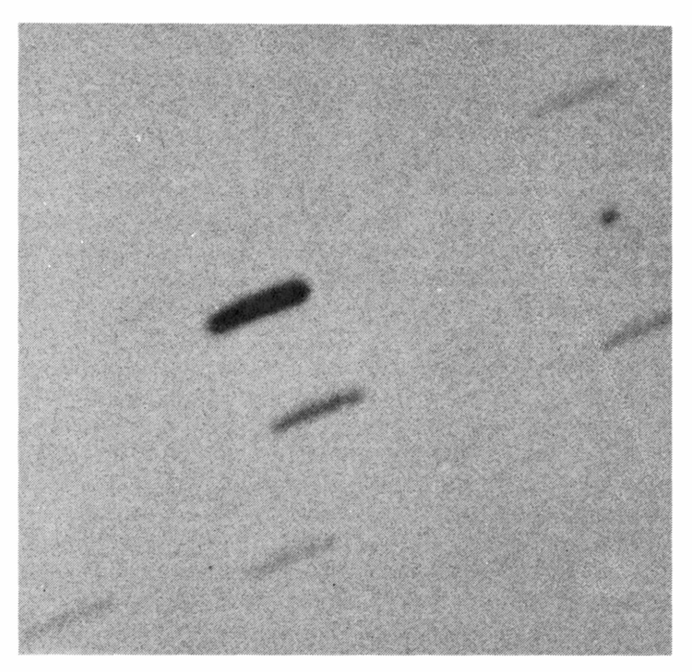
En el mismo año, por coincidencia, se descubrió otro pequeño asteroide, Amor, cuya órbita era semejante a la de Apolo excepto que su perihelio quedaba ligeramente fuera de la órbita terrestre a una distancia de 1,08 unidades astronómicas del Sol. (Una unidad astronómica, abreviada U.A., se define como la distancia media Sol-Tierra.) Posteriormente, se llegó a la conclusión de que las órbitas de objetos con perihelios tan próximos a 1 U.A. se transforman con frecuencia en órbitas que cortan a la terrestre. Los astrónomos llaman ahora objetos Apolo-Amor a la clase de cuerpos cuyos perihelios son menores que el valor, hasta cierto punto arbitrario, de 1,3 U.A.

Desde 1932 se han descubierto un total de 28 objetos Apolo, aproximadamente, que cortan a la órbita terrestre, y un número algo menor de objetos Amor. El "aproximadamente" es necesario porque en unos pocos casos las órbitas no están determinadas con la suficiente exactitud para poder afirmar con seguridad que estos objetos pertenezcan a la clase Apolo-Amor. Incluso con órbitas que parecían bien definidas, el destino de la mayoría de los primeros objetos Apolo-Amor descubiertos fue el de esfumarse rápidamente. Hubo que esperar a 1973 para volver a localizar al propio Apolo, y Adonis, el segundo de tales cuerpos (que se identificó en 1936), no se vio de nuevo hasta 40 años después; durante este período había cruzado la órbita terrestre más de 30 veces.

Hasta época reciente se había prestado escasa atención a los objetos Apolo-Amor, incluso por parte de la minoría de astrónomos especializados en estudios planetarios. Sólo gradualmente se llegó a la conclusión de que esos pequeños asteroides tienen una importancia para la ciencia terrestre y planetaria completamente desproporcionada a su tamaño y número. Los cuerpos del tipo Apolo-Amor han sido los principales productores de cráteres de diámetro superior a cinco kilómetros en la Tierra, la Luna, Mercurio, Venus y Marte (con la posible excepción de Marte). La densidad superficial de tales cráteres es, a su vez, el principal medio disponible para determinar la edad relativa de diferentes regiones de esos vecinos en el espacio. El recuento de cráteres hace posible correlacionar la geología de un planeta con la de otro.

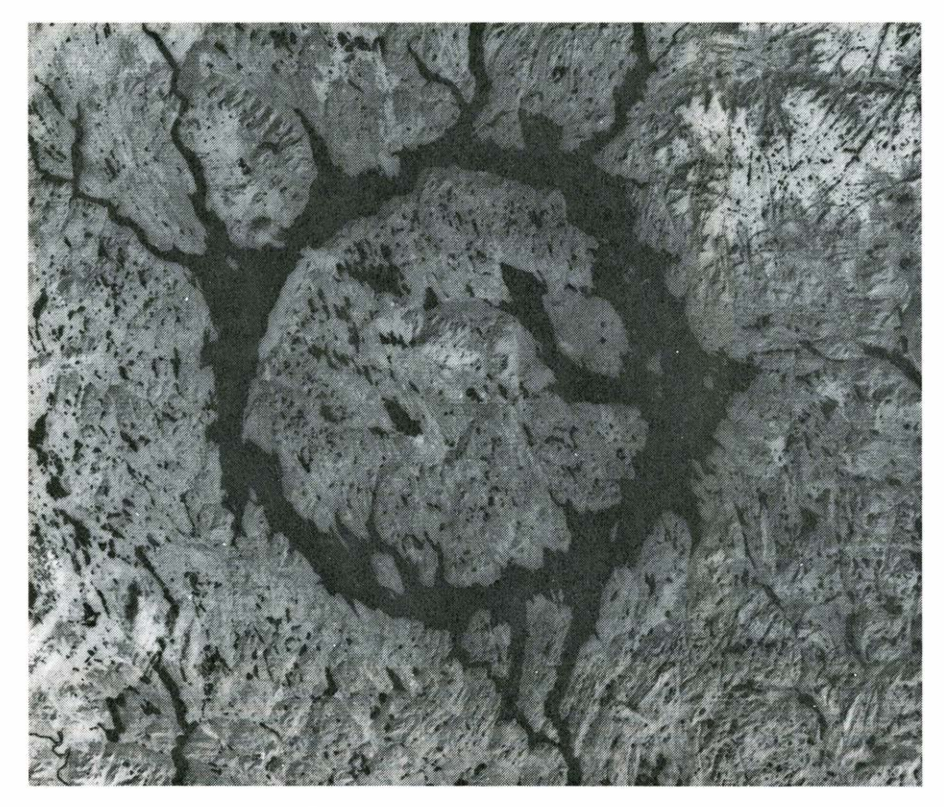
Una gran proporción, probablemente la mayoría, de los meteoritos que caen sobre la Tierra son fragmentos de objetos Apolo producidos en choques cuando las órbitas excéntricas de esos objetos les conducen a través del cinturón de asteroides, que se encuentra entre las órbitas de Marte y Júpiter. Los meteoritos son la fuente más importante de datos detallados sobre la primitiva historia del sistema solar. Para interpretar esos datos hay que conocer el origen de los meteoritos.

Se puede afirmar con cierta confianza que los objetos Apolo-Amor no son muestras al azar de cuerpos de algún modo expulsados del amplio cinturón de asteroides, que ocupa una región aproximadamente entre 2,1 y 3,5 U.A. del Sol. El primer asteroide, Ceres, fue descubierto en 1801 por el monje italiano Giuseppe Piazzi. Ceres, el mayor de los asteroides conocidos, mide unos 1000 kilómetros de diámetro y describe una órbita a unas 2,8 U.A. del Sol. En la actualidad se han catalogado más de 2000 asteroides, la mayoría de los cuales se encuentran más allá de la órbita de Marte (1,5 U.A.). Se estima que el cinturón de asteroides comprende un total de unos 400.000 objetos de diámetro superior a un kilómetro. Como veremos, algunos de esos cuerpos pueden convertirse en Apolos. Por otra parte, según una reciente y quizá sorprendente hipótesis, muchos cuerpos Apolo-Amor no son en realidad asteroides, sino restos de cometas desprovistos de sus gases, los

APOLO, el primer objeto Apolo conocido, fue descubierto en 1932, pero luego se perdió. Estas imágenes de negativo dejan constancia de su redescubrimiento 41 años más tarde, el 28 de marzo de 1973. Apolo es la tenue mancha redondeada que aparece en el centro de la foto de la izquierda. En la imagen de la derecha, tomada una hora después, aparece movido hacia el noroeste. Las estrellas brillantes, de la constelación de Virgo, que lo rodean tienen forma alargada porque el telescopio se movía a la velocidad angular prevista para los objetos Apolo durante los 25 minutos que duró cada exposición. Si se hubiesen mantenido fijas las imágenes de las estrellas, el rastro del tenue objeto Apolo, que se encontraba entonces a 100 millones de kilómetros de la Tierra, no hubiese dejado ninguna marca visible. Las fotografías del redescubrimiento fueron tomadas con el telescopio reflector de 155 centímetros del Harvard College Observatory por Richard F. McCrosky y Cheng-Yuan Shao, los cuales estaban preparados para realizar una búsqueda de varios meses de duración. Por una notable coincidencia Apolo ya se encontró en la primera fotografía. Se han descubierto 28 objetos Apolo.

2100 RA-SHALOM, el responsable del corto trazo que aparece en esta imagen de negativo, fue el tercer objeto Apolo que se descubrió el pasado año. Eleanor F. Helin, del Instituto de Tecnología de California, lo observó sobre una placa tomada el 10 de septiembre con un telescopio Schmidt de 46 centímetros en Monte Palomar. En realidad, este hallazgo era el redescubrimiento de un objeto fotografiado varios años antes y luego perdido. El número 2100 significa que es el que ocupa el lugar 2100 en el catálogo oficial de asteroides de todas las clases. A este objeto se le dio el nombre de "Ra", antiguo dios egipcio de la guerra, y "Shalom", palabra hebrea que significa paz, en el tiempo de las conversaciones tripartitas de Camp David.



CRATER GIGANTE EROSIONADO, al noreste de Quebec; se formó hace unos 210 millones de años a causa del impacto de un gran cuerpo similar a los Apolo. Para excavar un tal cráter, que tiene alrededor de 70 kilómetros de anchura, el asteroide debió tener un diámetro de aproximadamente tres kilómetros. La fotografía fue tomada por el Satélite para la Tecnología de los Recursos Terrestres (ERTS). La visibilidad del cráter se ve incrementada por el agua obscura del Lago Manicouagan, que ahora llena la depresión circular del cráter. En el centro de la formación aparece un saliente rocoso producido en el choque. En Canadá muchos cráteres estaban preservados bajo estratos de sedimentos que fueron arrastrados por la acción de los glaciares. Los objetos Apolo causaron grandes cráteres.



EL CRATER DEL METEORO, cerca de Flagstaff, en Arizona, mostrado en la misma escala que el cráter de arriba. El Cráter del Meteoro, apenas visible en el centro de la fotografía, mide 1300 metros de ancho. Parece haber sido formado hace unos 25.000 años por un meteorito férreo de menos de 100 metros de diámetro. El impacto liberó una energía equivalente a una explosión de 10 megatones.

objetos más primitivos que existen en el sistema solar.

En el pasado decenio ha aumentado grandemente el interés por los objetos Apolo-Amor. Cuando los astrónomos observan un cuerpo de movimiento rápido en sus placas fotográficas, comunican rápidamente su descubrimiento a la Oficina Central para Telegramas Astronómicos de la Unión Astronómica Internacional, así como a otros astrónomos. Se calculan y distribuyen efemérides indicando la posición, velocidad y dirección de movimiento, de manera que se puedan realizar observaciones de seguimiento. En algunos casos, el objeto recién descubierto puede ser reconocido en placas antiguas. Una vez que un cuerpo se ha identificado en varias placas, se puede generalmente calcular su órbita con precisión suficiente para garantizar su identificación en futuros encuentros.

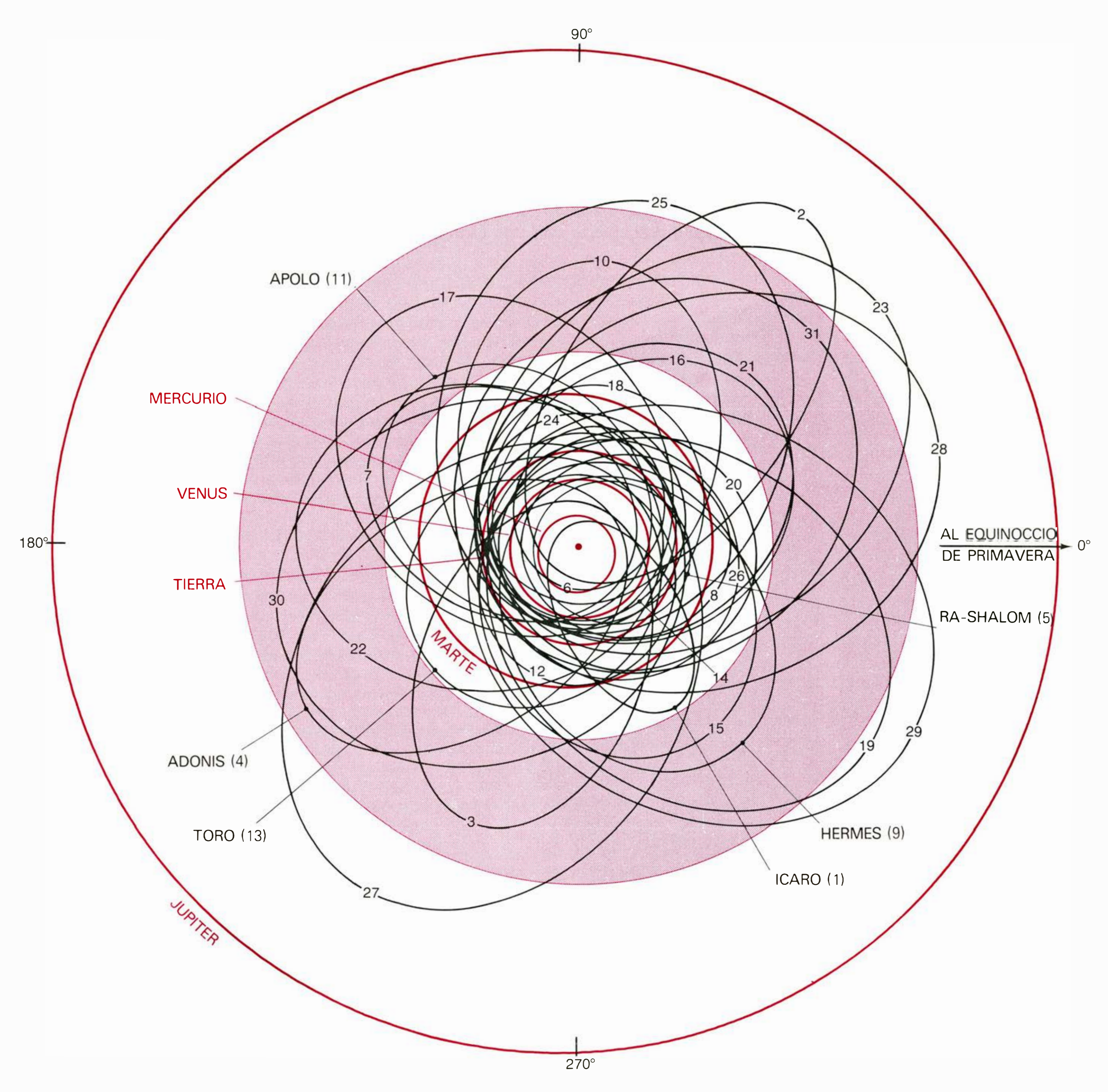
En 1973 comenzó el primer programa de observación dedicado a la búsqueda sistemática de objetos Apolo-Amor en el Observatorio de Monte Palomar a cargo de Eugene M. Shoemaker y Eleanor F. Helin, del Instituto de Tecnología de California. Trabajando con un telescopio Schmidt de gran campo y tamaño intermedio unas cuatro noches por mes, han descubierto cinco objetos Apolo y un objeto Amor, junto con cierto número de otros cuerpos interesantes, entre ellos un nuevo cometa.

Varios otros programas de observación que emplean telescopios Schmidt mayores, aun no estando expresamente dedicados a la búsqueda de objetos Apolo-Amor, han aportado contribuciones importantes. Charles Kowal, trabajando con el telescopio Schmidt de 1,2 metros en Monte Palomar, ha descubierto cinco objetos Apolo-Amor; un grupo que está realizando una exploración de todo el cielo en el Observatorio Austral Europeo en Chile ha descubierto tres.

En contraste con la situación de hace un decenio, casi la mitad de los objetos Apolo-Amor son asteroides completamente conocidos, designados por número. En estos días pocos de ellos se pierden irremediablemente. La mayor parte de los escasamente 40 objetos tienen órbitas fuertemente excéntricas que se extienden hacia fuera hasta el cinturón principal de asteroides y en media docena de casos incluso más allá de él. Sin embargo, al menos tres de los objetos tienen órbitas de excentricidad moderada, con un afelio (máxima distancia al Sol) dentro de la órbita de Marte.

Cualquier tratamiento cuantitativo del papel que los objetos Apolo-Amor juegan en la producción de cráteres en los planetas o como fuente de meteoritos requiere conocimiento de sus tamaños y abundancia. Su diámetro angular está muy por debajo de los límites de resolución de los telescopios, de modo que sus dimensiones se han de determinar indirectamente. Ello puede hacerse midiendo su albedo geométrico, es decir, la fracción que reflejan de la radiación solar que incide sobre ellos. Combinando el albedo de un objeto y su brillo medido con su distancia conocida a la Tierra y el Sol, se puede calcular su sección recta y, por tanto, suponiéndolo aproximadamente esférico se tiene idea de su tamaño.

El albedo se puede determinar de varias maneras. En régimen estacionario, la energía incidente que no es reflejada por un objeto debe ser absorbida y después reemitida, principalmente en longitudes de onda correspondientes al infrarojo. De aquí que midiendo el brillo de un objeto en la banda del infrarrojo y comparándolo con su brillo en longitudes de onda del espectro visible se puede determinar la reflectividad del objeto. Mediante medidas de laboratorio se ha demostrado también que el grado de polarización de la luz reflejada por una superficie depende conjuntamente del ángulo de reflexión y del albedo. Además, el "color" de un asteroide se puede de-



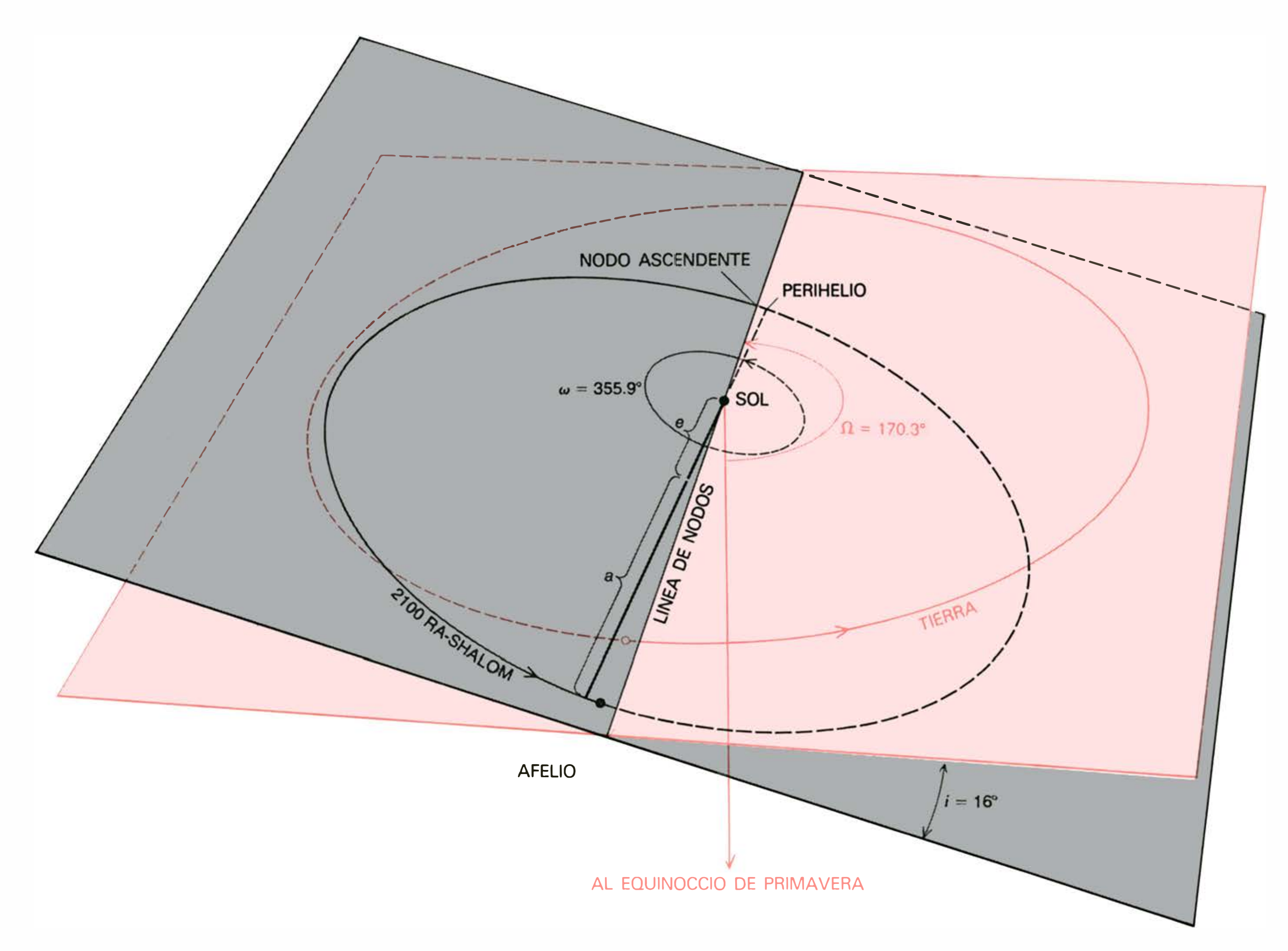
SE REPRESENTAN LAS ORBITAS de 31 objetos Apolo en relación con las órbitas de Mercurio, Venus, la Tierra, Marte y Júpiter. Están numeradas en orden de su distancia creciente al Sol en el perihelio, o punto más próximo al Sol. El asteroide que más se acerca al Sol es Icaro. Sólo tres de los objetos Apolo (Ra-Shalom, Nº 6 y Nº 14) ocupan órbitas que quedan dentro de la órbita de Marte. La mayoría cortan el cinturón principal de asteroides (*banda ancha de color*), que se estima incluye casi medio millón de objetos de diámetro superior a un kilómetro. Los fragmentos de objetos Apolo que han chocado con objetos en el cinturón de asteroides pueden proporcionar una gran parte de los meteoritos que caen en la Tierra.

terminar midiendo su brillo relativo a diferentes longitudes de onda en las regiones ultravioleta, visible e infrarrojo próximo del espectro. Tales medidas permiten con frecuencia al menos una identificación provisional de asteroides con conjuntos de minerales encontrados en meteoritos y en rocas lunares y terrestres. Mediciones de laboratorio del albedo de agrupaciones minerales semejantes pueden proporcionar otra comprobación sobre el albedo del asteroide.

Es fácil encontrar puntos débiles a cada uno de esos métodos para deducir el albedo de los asteroides, pero el hecho notable es que muchos de los albedos se han determinado por más de uno de los citados métodos y casi todos los resultados están en buena concordancia. La mayoría de las medidas se refieren a cuerpos típicos en el cinturón de asteroides. Solamente once cuerpos Apolo y Amor han sido hasta ahora objeto de medidas de albedo. Sin embargo, los albedos geométricos de esta pequeña muestra son análogos a los de los asteroides ordinarios, y varían desde 0,05 hasta 0,2. (Para comparación, el albedo geométrico de la Luna es 0,12.) Si se toma el albedo 0,16 como típico de los objetos Apolo-Amor, la mayoría de los descubiertos resultan tener un diámetro de entre uno y dos kilómetros. Con un albedo de 0,16, un objeto de alrededor de un kilómetro de diámetro tendrá una magnitud astronómica de + 18 cuando está situado a una distancia de 1 U.A. tanto de la Tierra como del Sol. La magnitud de un asteroide ajustada a estas condiciones patrón se llama su magnitud absoluta.

El número total de Apolos y Amores con una magnitud absoluta más brillante que + 18 (es decir, con un diámetro de más de un kilómetro) se puede estimar comparando el número realmente descubierto con una estima de lo completa que ha sido la búsqueda telescópica. El primer intento de tal clase fue realizado por Ernst J. Öpik, del Observatorio Armagh, en Irlanda, en 1963, cuando se conocían diez objetos Apolo. Öpik llegó a la conclusión de que debía haber al menos 43 Apolos y que "podía haber muchos más". Como se han descubierto ya 28 Apolos y el ritmo al que se descubren, ahora unos cuatro por año, está aumentando, debe realmente haber muchos más de 43.

Una estima más reciente, basada en el método de completitud de búsqueda y realizada por Shoemaker, llega al resultado de que el número total de objetos Apolo de magnitud más brillante que + 18 es 750 ± 300. En 1967, Fred L. Whipple, de la Universidad de Harvard, enfocó el problema de la estimación en forma bastante distinta, razonando que si el número total de Apolos era pequeño, sería muy probable que alguno de ellos fuese redescubierto por casuali-



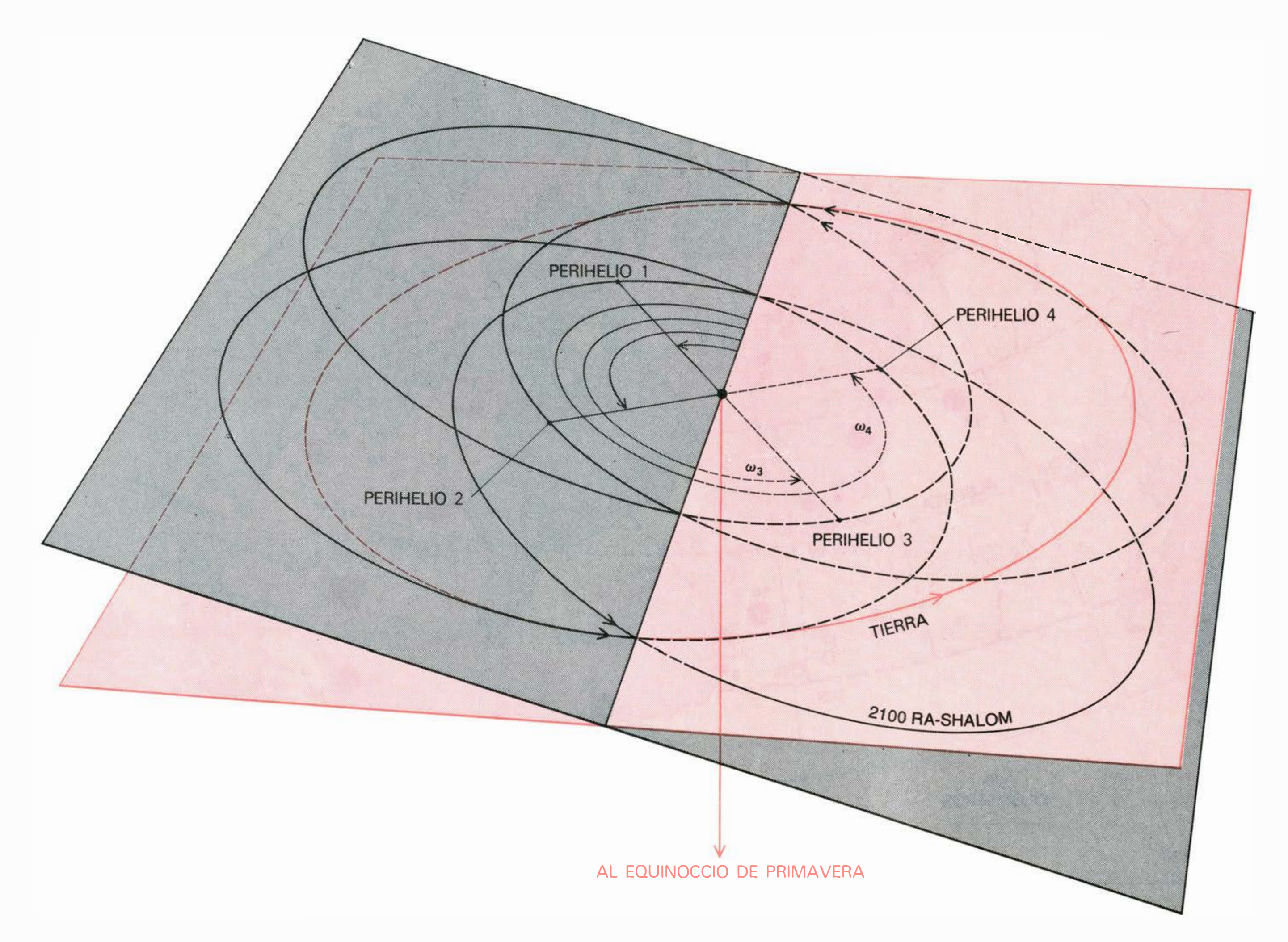
LA ORBITA DE LA TIERRA Y LA DE LOS OBJETOS APOLO pasan frecuentemente una por dentro de la otra como dos eslabones de una cadena. Sin embargo, como las órbitas de los objetos Apolo tienen una precesión rápida, su orientación con respecto a la órbita terrestre puede cambiar apreciablemente en unos pocos cientos de años. El plano orbital del 2100 Ra-Shalom está inclinado 16 grados (el valor *i*) respecto al plano orbital de la Tierra. Se muestran las posiciones relativas de los dos cuerpos en la noche del descubrimiento de Ra-Shalom, el 10 de septiembre. El tamaño y forma de la órbita elíptica de un asteroide vienen dados por la longitud del semieje mayor, *a*, y la excentricidad de la elipse, *e*, definida por *c*/*a* donde *c* es la distancia del Sol al centro de la elipse. La orientación del plano orbital del asteroide respecto al de la Tierra está dada por Ω, longitud del nodo ascendente, definida como la distancia angular medida hacia el este en el plano de la órbita terrestre, o eclíptica, desde el equinoccio de primavera al

dad. Hasta 1967 no se habían hecho redescubrimientos casuales. El álgebra de combinaciones y permutaciones se puede usar para establecer un probable límite inferior del número total de objetos Apolo. Whipple estimó al principio que había más de 50 y una repetición de este cálculo en 1973 dio un valor de 100. Sobre la base de datos más recientes, incluyendo el redescubrimiento casual en 1978 de un Apolo (Ra-Shalom), el método arguye poderosamente contra la existencia de menos de 200 y conduce a una estima que concuerda con la de Shoemaker.

El que un objeto Apolo choque o no con la Tierra es cuestión de suerte. En el momento presente, ninguno de los objetos Apolo conocidos sigue una trayectoria de choque con la Tierra. Sin embargo, todos los Apolo-Amor están bajo la continua influencia gravitatoria de planetas cercanos, particularmente Júpiter, que produce una precesión sobre las órbitas asteroidales. Como resultado de la precesión, el eje mayor de una órbita elíptica gira gradualmente en el espacio hasta completar 360 grados. Esto significa que todos los objetos con perihelio dentro de la órbita de la Tierra y afelio más allá de ella se encontrarán tarde o temprano en una órbita que corta a la de la Tierra. Por tanto, un objeto Apolo típico se encontrará en órbita que corta a la terrestre aproximadamente una vez cada 5000 años.

En la mayor parte de los casos, la Tierra se encontrará en algún otro punto de su órbita cuando el objeto Apolo se halle en el punto de intersección, de modo que los dos cuerpos no chocarán. Hay, sin embargo, una probabilidad calculable de que tanto la Tierra como el objeto Apolo se encuentren simultáneamente en el punto de intersección. Para cualquier objeto Apolo dado, la probabilidad de choque es sólo de alrededor de 5×10^{-9} por año, o una vez cada 200 millones de años. Suponiendo que haya entre 750 y 1000 objetos Apolo de diámetro superior a un kilómetro, resulta que aproximadamente cuatro deben chocar con la Tierra cada millón de años. Aunque tal frecuencia de choque es baja en los 5000 años de historia humana escrita, es alta en la escala de miles de millones de años del tiempo geológico.

Dentro de la historia humana con testimonio escrito, no se ha producido ninguna catástrofe resultante del choque de un Apolo. El más reciente encuentro conocido con un objeto asteroidal de tamaño apreciable tuvo lugar hace entre 25.000 y 50.000 años, cuando un cuerpo asteroidal férreo que probablemente no tenía más de 100 metros de diámetro excavó un cráter de algo más de un kilómetro de diámetro y casi 200 metros de profundidad, el Cráter del Meteoro, en Arizona. El cráter más jo-



punto en que el asteroide corta la eclíptica de sur a norte. La orientación de la elipse del asteroide en su plano orbital viene dada por ω, que es el ángulo entre el nodo ascendente y el perihelio medido en la dirección del movimiento del asteroide. Para el Ra-Shalom, *a* es de 0,83 unidades astronómicas (o 0,83 veces la distancia media de la Tierra al Sol), *e* vale 0,436, Ω es de 170,3 grados y ω 355,9 grados. Las perturbaciones causadas principalmente por la atracción gravitatoria del planeta Júpiter obligan al ángulo ω a girar una vuelta completa (360 grados) en unos 10.000 años. La configuración en hoja de trébol de la ilustración a la derecha muestra que para cuatro valores de ω separados por unos pocos miles de años, la órbita del asteroide cortará a la de la Tierra. Simultáneamente, y más o menos a la misma velocidad, la línea de los nodos sufrirá una precesión de una circunferencia completa (no representada en la ilustración). Aproximadamente una vez cada 250 millones de años chocarán un Apolo típico y la Tierra.

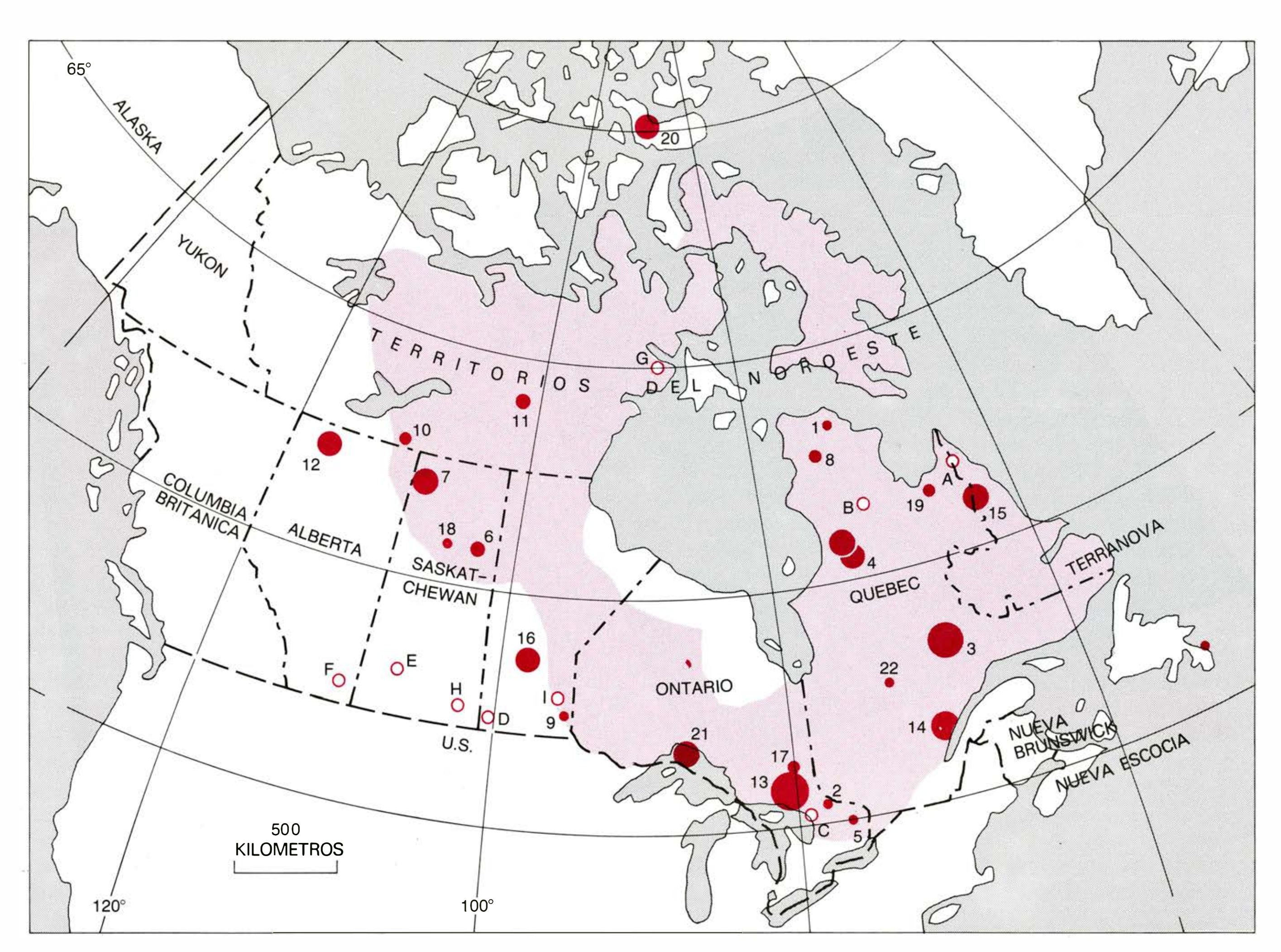
ven que se conoce con diámetro superior a 10 kilómetros es el que forma el lecho del Lago Bosumtwi en Ghana. El suceso que lo creó tuvo lugar hace 1,3 millones de años. Los cráteres tan pequeños como el del Meteoro son rápidamente erosionados. Sólo los mayores cráteres dejan rastro después de 500.000 años. Si no fuese por los efectos erosivos de la intemperie, procesos geológicos y organismos vivientes, la Tierra tendría un aspecto tan densamente salpicado de cráteres como la Luna.

Cuando tienen lugar esos choques, raros pero estadísticamente inevitables, las consecuencias son dramáticas. En el sistema de referencia de la Tierra, la energía cinética de un objeto Apolo de un kilómetro de diámetro y densidad típica (3,5 gramos por centímetro cúbico) moviéndose en una órbita típica, es aproximadamente de 4×10^{27} erg. Esto equivale a 100.000 veces la energía liberada por la detonación de una bomba nuclear de un megatón. Toda esta energía debe disiparse de una manera u otra al chocar con la Tierra. Sólo se necesita una pequeña fracción de la energía total para vaporizar el cuerpo incidente, de modo que la mayor parte de la energía se consumirá en pulverizar el material de la superficie de la Tierra y expulsarlo a gran velocidad. De aquí que el impacto excavará un cráter similar a los que dominan los paisajes de la Luna, Marte y Mercurio.

No hay datos directos sobre las dimensiones de los cráteres producidos por explosiones de 10^5 megatones. Se puede, sin embargo, extrapolar relaciones empíricas para calcular las dimensiones basadas en explosiones mucho menores y llevar a cabo en un ordenador simulacros de explosiones de magnitud asteroidal. Tales cálculos indican que un impacto que libere 4×10^{27} erg de energía creará un cráter de 22 kilómetros de diámetro.

Si se acepta la estima de que la Tierra sufre, cada millón de años, aproximadamente cuatro impactos de objetos capaces de producir cráteres de 20 kilómetros de diámetro, cabe preguntarse ¿dónde están las pruebas? Se podría esperar que la Tierra estuviera abundantemente salpicada de cráteres. Resulta que las pruebas existen, pero hasta hace relativamente poco tiempo habían escapado a nuestra atención.

Una inspección detallada de las fotografías aéreas descubre muchas configuraciones circulares, frecuentemente muy poco marcadas, sobre la totalidad de la superficie terrestre. Un estudio geológico detallado ha demostrado que cierto



CASI LA MITAD DE LOS ANTIGUOS CRATERES reconocidos sin duda se han encontrado en el Canadá central y oriental, región que representa aproximadamente el 1 por ciento del área continental del planeta. Los círculos llenos muestran la posición y tamaño relativo de 23 cráteres de impacto bien establecidos, numerados en el orden de su descubrimiento; los círculos no llenos identifican posibles estructuras de impacto. El cráter más joven conocido (N.º 1) tiene menos de cinco millones de años; el más antiguo (N.º 13, cerca de Sudbury, Ontario) tiene unos 1800 millones de años. Este último es uno de los cráteres que se pueden atribuir al período precámbrico, que terminó hace 600 millones de años. El cráter N.º 3 es la formación Manicouagan de la fotografía superior de la página 10. Richard Grieve, del Departamento Canadiense de Energía, Minas y Recursos, ha hecho un recuento de los cráteres antiguos bien establecidos en Norteamérica y Europa; sobre la base de su recuento, se puede estimar que en los pasados 600 millones de años, 1500 objetos Apolo de diámetro igual o superior a un kilómetro han chocado con la Tierra, de los cuales presumiblemente un 70 por ciento lo hicieron en los océanos. El mapa se basa en una compilación hecha por la Sección de Física de la Tierra del Departamento de Energía.

número de ellas son los restos de cráteres de impacto que en su origen eran de diámetro superior a 20 kilómetros. Su origen queda probado por el descubrimiento de estructuras geológicas que uno esperaría encontrar en cráteres de explosión profundamente erosionados. Hay también claras pruebas de que las rocas en la región han estado sometidas a extremada compresión. Las pruebas incluyen la presencia de rocas ígneas que han recristalizado después de haberse fundido por un súbito impacto, y la presencia de formas de cuarzo, tales como la coesita y la estishovita, que pueden formarse únicamente gracias a la concurrencia de altas presiones y altas temperaturas. Entre las pruebas macroscópicas se cuenta la presencia de "conos de astillamiento", estructuras de cuarzita que se extienden hacia fuera y hacia abajo alejándose de la dirección del impacto. En algunos casos, en los productos expulsados del cráter se puede identificar la "huella" química de un cuerpo extraterrestre que chocó. Sin embargo, sólo en raras ocasiones se hallan fragmentos del cuerpo incidente: éste se vaporiza como consecuencia de las temperaturas que se alcanzan en la explosión.



CONOS DE RUPTURA. Se trata de rocas que se han fracturado por impacto según un modelo cónico; se han encontrado en la mayoría de los emplazamientos de cráteres que están bien establecidos. Estos conos se hallaron en un afloramiento en la orilla sur del lago Keeley, cerca de la depresión de Sudbury, resto de un cráter con diámetro original de 140 kilómetros. Es la estructura 13 del mapa anterior.

Tan claras pruebas de su origen por choque han conducido en la actualidad a la identificación de más de 50 estructuras debidas a impactos, la mayor parte de ellas, de manera no sorprendente, en regiones de la Tierra habitadas por geólogos interesados en tales estructuras. Aproximadamente la mitad de los cráteres bien establecidos se han hallado en el Canadá central y sudoriental, que forma sólo aproximadamente el 1 por ciento de la superficie terrestre. Como no hay manera natural mediante la cual los cuerpos extraterretres puedan ser canalizados hacia una pequeña porción de la superficie terrestre, la abundancia de cráteres canadienses solamente se puede atribuir a la diligencia del programa de investigación organizado por el Departamento Canadiense de Energía, Minas y Recursos de Ottawa.

La mayor parte de los cráteres estudiados son geológicamente jóvenes y no tienen más de unos pocos cientos de millones de años. Solamente se conocen dos cráteres precámbricos (uno cerca de Sudbury, en Ontario, y el otro cerca de Vredefort, Africa del Sur) aun cuando la época precámbrica abarca más del 80 por ciento de la historia de la Tierra. A menos que un cráter sea desusadamente grande, todo rastro de su presencia habrá sido borrado en menos de 600 millones de años. Además, los cráteres producidos por cuerpos asteroidales que caigan en los océanos, que cubren más del 70 por ciento de la superficie de la Tierra, son, comprensiblemente, difíciles de identificar. Teniendo todo ello en cuenta, se considera ahora que la velocidad de formación de cráteres por unidad de superficie en la Tierra es al menos tan grande como lo es en la Luna.

Muchos de los cráteres terrestres identificados tienen un diámetro no superior a los 100 metros; por tanto, caen fuera del margen de tamaños de cráteres que pueden ser excavados por los mayores meteoritos conocidos. Se pueden llamar con propiedad cráteres de meteoritos. Sin embargo, aproximadamente la mitad de los cráteres terrestres conocidos tienen un diámetro superior a los 10 kilómetros; por tanto, deben haber sido producidos por cuerpos mucho mayores que ninguno de los meteoritos hasta ahora encontrados. En otras palabras, deben haber sido producidos por el impacto de cuerpos de tamaño comparable al de los objetos Apolo visibles telescópicamente. Una estima plausible, basada en la distribución de cráteres conocidos, calcula que unos 1500 objetos semejantes a los Apolo han chocado con la Tierra desde el final del período precámbrico, hace 600 millones de años. Considerando la presente incertidumbre en el número total, tamaño y velocidad de los objetos Apolo, esta estima concuerda razonablemente bien con la dada arriba de cuatro choques por cada millón de años.

¿Cómo se puede estar seguro de que es correcto identificar los grandes cráteres de la Tierra y la Luna con los objetos Apolo, que sólo recientemente han sido descubiertos? La respuesta es que los objetos capaces de producir grandes cráteres deben ser suficientemente grandes para poder verse con telescopios. Sólo hay dos clases de objetos astronómicamente observables que cortan la órbita de la Tierra: los objetos Apolo y los cometas. Al contrario que los asteroides, el núcleo sólido de los cometas contiene hielos u otras substancias que, cuando el cometa se acerca al Sol, se volatilizan y producen una coma nebulosa, o cabeza, y frecuentemente también una cola. Las comas suelen expandirse hasta un diámetro de miles de kilómetros. El diámetro del núcleo dentro del coma se puede estimar comparando los valores observados y calculados de la cantidad de gas que se espera sea emitida en función de la distancia del cometa al Sol. Una indicación más directa respecto al tamaño es el brillo del núcleo de un cometa cuando se encuentra tan lejos del Sol que el coma que lo enturbia no existe. Tales estimas indican que los núcleos de los cometas están en la misma categoría de tamaños que la deducida para los objetos Apolo: de un kilómetro a 10 kilómetros.

Puesto que tanto los objetos Apolo como los cometas son capaces de chocar con la Tierra y la Luna, y dejar cráteres en ellas, ¿cómo puede estimarse el ritmo relativo de bombardeo para los dos tipos de objetos? Si se tiene en cuenta únicamente estimas del tamaño y número de los objetos Apolo, parece que ellos por sí solos son suficientemente abundantes para responder del número de cráteres grandes producidos en la Luna desde el final de la inundación de los "mares" por magmas basálticos, hace unos tres mil millones de años. Como hemos visto, la abundancia estimada de Apolos concuerda también con el número estimado de grandes cráteres producidos en la Tierra desde los tiempos precámbricos.

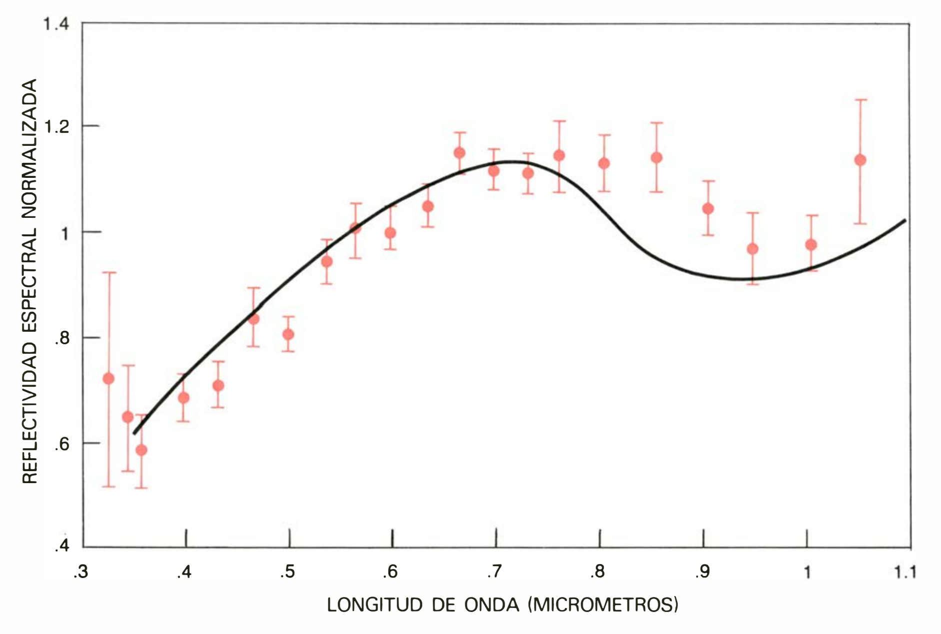
Puede demostrarse también que si el núcleo de los cometas no es mayor que los objetos Apolo, los cometas sólo pueden haber producido un pequeño porcentaje de los cráteres mayores observados en la Luna. La fracción, sin embargo, podría ser mucho mayor si los cometas observados fueran, en realidad, mayores de lo que se ha estimado y si, como resultado, hubiese muchos cometas no observados de diámetro superior a un kilómetro. En tal caso, los impactos debidos a cometas podrían dar cuenta de hasta el 35 por ciento de los cráteres lunares y terrestres más grandes. Recientemente, A. H. Delsemme, de la Universidad de Toledo (Ohio) y L'ubor Kresák, del Instituto Astronómico de Bratislava en Checoslovaquia, han presentado, cada uno por su parte, pruebas de que el número de cometas débilmente luminosos es en realidad más bajo de lo que podría esperarse, lo que arguye a favor de que la contribución cometaria a la producción de cráteres sea verdaderamente pequeña.

Habiendo trazado la distinción entre objetos Apolo y cometas, voy ahora a hacerla más borrosa mencionando una posibilidad sobre la que luego volveré en más detalle. Como sugirió en un comienzo Öpik, muchos objetos Apolo pueden ser cometas "extintos", esto es, residuos no volátiles de núcleos de cometas antaño activos que han sido completamente desprovistos de sus gases por repetidas pasadas cerca del Sol. Si éste es el caso, muchos cráteres terrestres y lunares son cometarios en el sentido de que el cometa responsable de ellos se había previamente metamorfoseado en un objeto Apolo y, por tanto, está ya incluido en el cálculo de velocidad de producción de cráteres por Apolos.

¿De qué manera, si es que hay alguna, son los objetos Apolo capaces de formar cráteres relacionados con las miriadas de cuerpos menores, de tamaños cambiantes desde grandes meteoros hasta partículas de unos micrometros, que continuamente bombardean la atmósfera terrestre? Parece probable que esos cuerpos menores deriven de fuentes de diversas clases. Se puede decir en seguida, no obstante, que una elevada proporción de los meteoros observados (a diferencia de los meteoritos) no están relacionados con objetos de apariencia asteroidal sino que derivan de cometas observables. Los meteoros cometarios son los que se ven en los chubascos periódicos de meteoros, tales como las Perseidas, que siguen la órbita del Cometa Swift-Tuttle y las Dracónidas, asociadas con el cometa de período corto Giacobini-Zinner. También hay meteoros asociados a cometas activos de los que se ha observado cómo se desintegraban y desaparecían, por ejemplo el cometa Biela, que dio lugar a las Andromédidas. Es por tanto plausible que al menos algunos de los chubascos meteóricos aparentemente no relacionados con cometas conocidos representen en realidad los restos de cometas que se desintegraron en los últimos pocos miles de años, cuando aún estaban en plena actividad.

Por otra parte, una gran proporción de los meteoros observados y estrellas fugaces no tienen relación obvia con los chubascos de meteoros. Estos incluyen los meteoritos que sobreviven al paso a través de la atmósfera terrestre. Resulta muy improbable que ninguno de los meteoritos que actualmente se encuentran en colecciones de museos llegase como parte de un chubasco, aun cuando algunos podrían ser los restos de chorros cometarios dispersados. Se puede determinar el tamaño que los cuerpos meteoríticos tenían en órbita midiendo el crecimiento en ellos de productos de reacciones nucleares desencadenadas por los rayos cósmicos galácticos. El resultado es que casi todos los meteoritos han tenido un tamaño similar al actual de menos de un metro durante millones de años. Incluso si los meteoritos se originaron en chorros, como restos de cometas, habrían sido dispersados en órbitas al azar por interacción gravitatoria con los planetas a lo largo de unos pocos miles de años.

Es cierto, sin embargo, que no todos los meteoros son restos de cometas. Por ejemplo, puede verse que una proporción pequeña pero apreciable de meteoros, presumiblemente constituidos por un material lo bastante resistente para sobrevivir al paso a través de la atmósfera terrestre, debe representar fragmentos producidos por choque entre asteroides ordinarios en el cinturón principal de asteroides entre las órbitas de Marte y Júpiter. Como resultado de perturbaciones gravitatorias inducidas primordialmente por Júpiter, algunos de los fragmentos serán "empujados" a órbitas fuertemente excéntricas que cortan



REFLECTIVIDAD ESPECTRAL DEL OBJETO APOLO designado 1865 Toro (*barras verticales*). Dicha reflectividad se ajusta bien a la reflectividad de una condrita típica, el tipo más común de litometeoritos (*curva de trazo continuo*). El ancho mínimo relativo en la curva espectral del meteorito en la región de 0,95 micrometros corresponde a una fuerte banda de absorción del mineral piroxeno. Presumiblemente, el piroxeno es un componente importante de 1685 Toro, único Apolo estudiado en detalle espectral. Las medidas fueron realizadas por Chapman, del Instituto de Ciencia Planetaria, T. B. McCord y M. J. Gaffey, de la Universidad de Hawai, y C. Pieters, del Centro Espacial Johnson.

a la órbita terrestre. Cualquiera de esos fragmentos que por azar entre en la atmósfera terrestre se pondrá incandescente y se verá en forma de meteoro; los que sobreviven y alcanzan el suelo son los meteoritos. Evidentemente, no hay posibilidad de identificar de un modo directo tales objetos con su origen en el cinturón de asteroides por medio de sus órbitas.

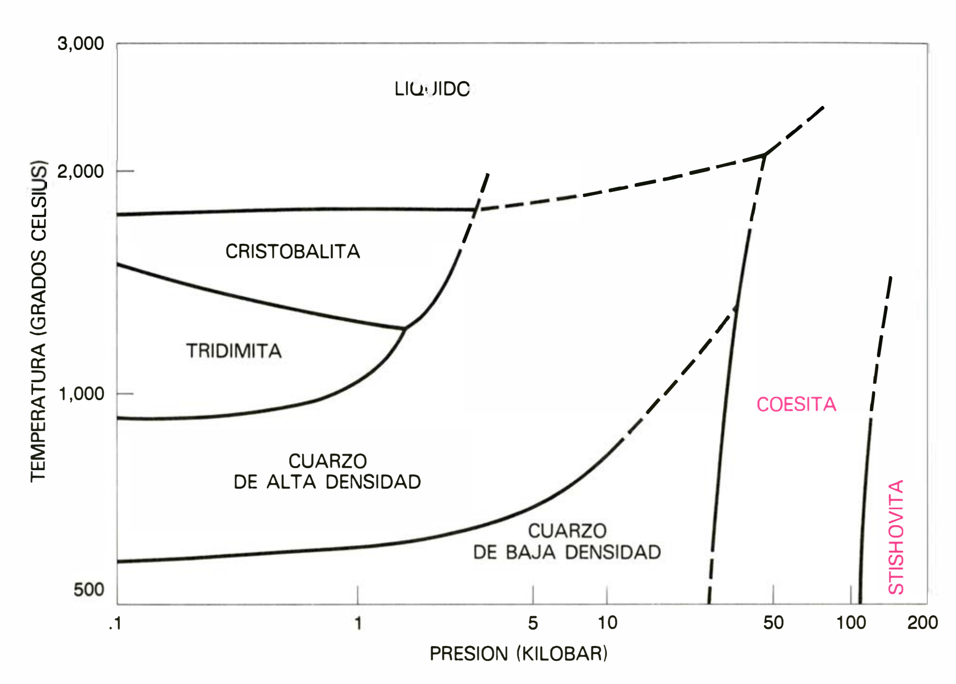
Cualitativamente se puede presentar una buena argumentación para creer que algunos meteoros y meteoritos son fragmentos directos de objetos Apolo. Al menos 19 de los objetos Apolo tienen órbitas que bien penetran en el cinturón principal de asteroides o bien se extienden más allá de él. De aquí que inevitablemente llegarán a chocar con el gran volumen de residuos concentrados en esa parte del sistema solar. Además de varios cientos de grandes cuerpos con diámetros entre 25 y 1000 kilómetros, se estima que la región contiene al menos 10^{10} objetos cuyo tamaño varía de un kilómetro a unos pocos metros. Los choques entre esos numerosos asteroides pequeños y los objetos Apolo deben ser frecuentes.

En un encuentro típico, el pequeño objeto producirá un cráter en el Apolo, que tiene un tamaño mayor. El choque imprimirá una velocidad de unos 50 metros por segundo al material expulsado, que se dispersará en el espacio porque la velocidad de escape respecto a un cuerpo de un kilómetro de diámetro es muy pequeña, de menos de un metro por segundo. La velocidad adicional de 50 metros por segundo impresa a los restos no será suficiente, sin embargo, para cambiar su órbita apreciablemente respecto a la del Apolo mismo, que se mueve alrededor del Sol con una velocidad de unos 20 kilómetros por segundo. En consecuencia, los fragmentos resultantes del choque tendrán, en un período de unos pocos millones de años, una apreciable probabilidad de penetrar en la atmósfera terrestre.

Si un objeto Apolo choca con un asteroide de diámetro de hasta 100 metros, hay una alta probabilidad de que el propio Apolo se fracture. El enjambre de restos tendrá luego la misma probabilidad que tenía el Apolo de chocar con la Tierra. Sin embargo, el intervalo de tiempo que probablemente se necesitará para el choque con la Tierra será mucho mayor que el tiempo necesario para dispersar la corriente de fragmentos a órbitas propias, independientes y esencialmente aleatorias. Como consecuencia, los meteoros y meteoritos producidos por la formación de cráteres en un objeto Apolo o por la fractura del mismo es mucho más probable que lleguen a la Tierra individual y esporádicamente que no que lo hagan como miembros de una corriente de fragmentos. Esto, por supuesto, concuerda con las observaciones.

Yo he tratado de cuantificar el argumento cualitativo que acabo de presentar a fin de ver qué proporción del material meteórico y meteorítico que se estima bombardea anualmente la Tierra podría ser explicado como consecuencia de sucesos del tipo choque contra objetos Apolo en el cinturón principal de asteroides. Mi estima es que en dominio de tamaños entre 100 gramos y 1000 kilogramos (una tonelada métrica) deben penetrar en la atmósfera terrestre unas 400 toneladas métricas de fragmentos Apolo cada año. Si los fragmentos tienen las propiedades mecánicas de los litometeoritos típicos, aproximadamente 100 de las 400 toneladas deben sobrevivir al paso a través de la atmósfera terrestre y ser recuperables en forma de meteoritos. Aunque este cálculo es inseguro en un factor de al menos 10, el valor de 100 toneladas es muy próximo a las estimas actuales del ritmo de impacto de meteoritos determinado a partir de la recogida real de los mismos, mediante la red de cámaras de vigilancia de meteoritos brillantes y mediante el ritmo de impactos detectado por los sismómetros colocados en la Luna.

A menos que el material Apolo sea excepcionalmente frágil y susceptible de destrucción en la atmósfera terrestre, parece virtualmente cierto que una elevada proporción de los 2000 meteoritos en las diversas colecciones sean fragmentos Apolo, y es posible que lo sean la casi totalidad de ellos. La competencia de otras posibles fuentes no es fuerte. Ya he razonado que los meteoritos recuperados no pertenecen a chorros asociados con cometas activos. Las únicas fuentes distintas que quedan llevan consigo procesos capaces de perturbar los fragmentos procedentes de choques entre cuerpos en el cinturón principal de asteroides, llevándolos a órbitas que corten la de la Tierra mediante interacciones gravitatorias resonantes con los planetas vecinos, principalmente Júpiter. La cosecha de meteoritos procedentes de tales mecanismos es difícil de calcular, pero las estimas más razonables sugieren que suministrarían sólo alrededor de una décima parte del material que bombardea la Tierra, en comparación con el que procede de la fragmentación de objetos Apolo.



LA PRESENCIA DE FORMAS DE CUARZO de alta densidad (coesita y stishovita) es una clara indicación de que las rocas terrestres han estado sometidas a altas presiones por el choque de un objeto de alta velocidad. La sílice tiene cierto número de formas cristalinas que son estables en diferentes dominios de presión y temperatura; las dos formas del cuarzo son las más corrientes. Las formas de alta temperatura tridimita y cristobalita han sido metastablemente preservadas por enfriamiento rápido en ciertas rocas volcánicas. La coesita y la stishovita, que se pueden formar solamente a presiones por encima de 30 kilobar se han encontrado en la vecindad de muchos cráteres de impacto terrestre.

Si aceptamos el hecho de que una gran cantidad, y quizá la mayor parte, de los meteoritos se han originado a partir de objetos Apolo, ¿qué pueden decirnos los meteoritos acerca de la naturaleza de tales objetos? Aproximadamente el 85 por ciento de todos los meteoritos hallados pertenecen a la clase de objetos pé-

treos conocidos por condritas. Las condritas se llaman algunas veces meteoritos no diferenciados porque su contenido de elementos menos volátiles se aproxima mucho a las proporciones de abundancia de elementos en el Sol y en el sistema solar en conjunto. El estudio de las condritas ha suministrado la mayor parte de los datos sobre la abundancia general de los elementos y sus isótopos, la edad del sistema solar y las condiciones en el primitivo sistema solar.

Clark R. Chapman, del Instituto de Ciencia Planetaria de Tucson, Thomas B. McCord y Michael J. Gaffey, de la Universidad de Harvard, y Carle Pieters, del Centro Espacial Johnson en Houston, midiendo la reflectancia espectral del 1685 Toro, primer cuerpo de los que cruzan la órbita terrestre que fue examinado de esta forma, suministraron un indicio que apoya la relación entre las condritas y los objetos Apolo. La reflectividad media del objeto Apolo en la región espectral que comprende desde la longitud de onda de 0,3 micrometros en el ultravioleta hasta la de 1,1 micrometros en el infrarrojo sigue muy de cerca la reflectividad de las condritas típicas. Tras subir hasta un máximo a 0,7 micrometros, ambos espectros muestran una caída en la vecindad de 0,95 micrometros antes de subir de nuevo. En el caso de la condrita, el descenso corresponde a una banda de absorción prominente atribuible al piroxeno, mineral

	OBJETO	DISTANCIA DEL PERIHELIO (UNIDADES ASTRONOMICAS)	DISTANCIA DEL AFELIO (UNIDADES ASTRONOMICAS)	INCLINACION ORBITAL (GRADOS)	DIAMETRO ESTIMADO (KILOMETROS)	AÑO DEL DESCUBRIMIENTO
1	1566 ICARO	0.19	1.97	23	1.0	1949
2	1978 SB	0.35	4.11	12	8.0	1978
3	1974 MA	0.42	3.09	38	4.0	1974
4	2101 ADONIS	0.44	3.30	1	0.8	1936
5	2100 RA-SHALOM	0.47	1.20	16	3.0	1978
6	1976 UA	0.47	1.22	6	0.2	1976
7	1864 DEDALO	0.56	2.36	22	2.0	1971
8	1865 CERBERO	0.58	1.58	16	1.0	1971
9	HERMES	0.62	2.66	6	1.0	1937
10	1981 MIDAS	0.62	2.93	40	1.0	1973
11	1862 APOLO	0.65	2.29	6	2.0	1932
12	2063 BACO	0.70	1.45	9	1.0	1977
13	1685 TORO	0.77	1.96	9	4.0	1948
14	2062 ATEN	0.79	1.14	19	1.0	1976
15	1977 HA	0.79	2.40	23	0.6	1977
16	PL-6743	0.82	2.42	7	0.4	1960
17	1976 WA	0.82	3.03	23	2.0	1976
18	1620 GEOGRAFO	0.83	1.66	13	2.0	1951
19	1947 XC	0.83	3.67	1	2.0(?)	1947
20	1959 LM	0.83	1.84	3	6.0(?)	1959
21	1950 DA	0.84	2.53	12	2.0	1950
22	1866 SISIFO	0.87	2.92	41	5.0	1972
23	1973 NA	0.88	3.98	68	3.0	1973
24	1978 CA	0.88	1.37	26	2.0	1978
25	1863 ANTINOO	0.89	3.63	18	2.0	1948
26	1975 YA	0.91	1.69	61	1.0	1975
27	PL-6344	0.94	4.21	5	0.4	1960
28	1978 DA	1.02	3.92	16	2.0	1978
29	1915 QUETZALCOATIL	1.05	3.99	21	1.0	1953
30	1917 CUYO	1.06	3.23	24	2.0	1966
31	1580 BETULIA	1.12	3.27	52	6.0	1950

VEINTIOCHO CUERPOS APOLO se conocen en la actualidad. Hay tres más, los tres últimos en la tabla, cuyas órbitas no cortan a la de la Tierra en el momento presente, pero llegarán a cortarla como resultado de perturbaciones causadas por Júpiter. 16 de los 31 se designan por un número, indicando que sus elementos orbitales se conocen con precisión suficiente para considerarlos incluidos en el catálogo de asteroides. La mayor parte de los restantes se identifican por el año del descubrimiento seguido de dos letras; la primera indica la quincena del año (empezando con la A para el 1 a 15 de enero) en la cual se realizó el descubrimiento, y la segunda el orden del descubrimiento en ese período. Dos objetos se designan PL, que significa exploración Palomar-Leiden. Ra-Shalom se designó originariamente 1978 RA, significando que era el primer objeto (A) que se descubría en la primera quincena de septiembre (R) de 1978. Los descubridores de un asteroide tienen el privilegio de darle nombre, honor que rige en otras ciencias.

que consta principalmente de silicatos de magnesio y hierro. Se puede inferir que el piroxeno es también abundante en el 1685 Toro.

Mediciones de albedo de otros Apolos apoyan la idea de que son condríticos. Son de especial interés tres Apolos de albedo muy bajo, incluyendo los dos últimos que se han descubierto: Ra-Shalom (cuyo nombre expresa la esperanza de paz universal) y 1978 SB, recién descubierto en el Observatorio Astrofísico de Crimea. Los albedos de esos dos objetos parecen ser suficientemente bajos para asemejarse a los de las condritas carbonáceas, que son el tipo químicamente más indiferenciado de meteoritos.

Estamos ahora en condiciones de preguntar: ¿de dónde vienen los objetos Apolo? La primera idea, la de que hayan existido desde que se formó el sistema solar hace 4500 millones de años, puede demostrarse fácilmente que es equivocada. Ya he indicado que la probabilidad de que un Apolo choque con la Tierra es aproximadamente de 5×10^{-9} por año, y casi la misma probabilidad se aplica cuando el choque se refiere a Venus. Por tanto, después de varios cientos de millones de años debería haber muy pocos supervivientes de la población inicial de objetos cuyas órbitas cortan a la terrestre. A pesar de todo, ¿podría la actual familia de Apolos, estimada en 750, representar los restos de una población que inicialmente era muy grande? La respuesta es negativa. Si hubiera existido una gran población, el registro de cráteres en la superficie de la Luna lo revelaría. El flujo de proyectiles que producen cráteres se ha mantenido casi constante durante los últimos tres mil millones de años. Si acaso, hay alguna indicación de que la velocidad de producción de cráteres ha aumentado ligeramente durante los últimos 600 millones de años.

Por tanto, queda claro que los Apolos están siendo introducidos en el sistema solar interior a un ritmo sensiblemente constante. En números redondos, alrededor de 15 nuevos Apolos de diámetro superior a un kilómetro deben proveerse cada millón de años para equilibrar la velocidad de desaparición estimada que supone cuatro choques con la Tierra y tres con Venus cada millón de años. Posiblemente, en el mismo período desaparecerá un Apolo por choque con algún otro cuerpo (Mercurio, la Luna o Marte) y los otros serán impulsados a órbitas hiperbólicas y, por tanto, expulsados completamente del sistema solar. Parece que haya solamente dos fuentes para los 15 Apolos necesarios. O bien son pequeños asteroides del cinturón principal de asteroides que de alguna manera han sido lanzados a órbitas que corten la órbita de la Tierra, o bien son los restos, desprovistos de sus gases, de cometas anteriormente activos. En la actualidad se comprende, al menos cualitativamente, el proceso mediante el cual ambas de esas fuentes podrían proporcionar el número necesario de objetos Apolo.

Los pequeños asteroides de tamaño semejante al de los objetos Apolo están siendo continuamente producidos en el cinturón principal de asteroides como fragmentos resultantes de choques entre cuerpos más grandes. Los pequeños asteroides se destruirán entre sí por choques, dando como resultado un número de ellos poco variable. La población estacionaria de asteroides de tamaño superior a un kilómetro en el cinturón principal es de unos 400.000. Se estima que se están formando y destruyendo por choques al ritmo de unos 400 por cada millón de años. Si existiese una manera de situar alrededor del 4 por ciento de los nuevamente formados en órbitas que corten a la de la Tierra, ello bastaría para mantener la población de Apolos.

Es muy improbable, sin embargo, que aun esta pequeña fracción de objetos pueda ser expulsada del cinturón principal de asteroides y lanzada a órbitas que corten a la de la Tierra como consecuencia de los choques que los originaron. La variación de velocidad necesaria para transferir materia de una órbita típica de asteroide a una órbita que corte la de la Tierra es de unos seis kilómetros por segundo. La energía cinética correspondiente es superior a 10^{11} erg por gramo. La energía requerida para fragmentar un cuerpo del tamaño de los Apolo no es superior a unos 10^{8} erg por gramo. Por consiguiente, cualquier choque capaz de suministrar la energía cinética necesaria para desviar un objeto del tamaño de los Apolo a un órbita que corte la de la Tierra desintegraría muy probablemente el objeto en fragmentos pequeños. Además, la experiencia con cráteres de explosión en la Tierra muestra que sólo una minúscula proporción del material expulsado alcanza velocidades de seis kilómetros por segundo; la mayor parte de él no llega a alcanzar los 200 metros por segundo. Como resultado, aun cuando un fragmento grande pudiera ocasionalmente sobrevivir a la expulsión a seis kilómetros por segundo, el número de tales fragmentos sería muy pequeño.

Es necesario, pues, buscar mecanismos más suaves capaces de realizar el cambio a una órbita que cruce la de la Tierra. Se conocen varios, y todos tienen la propiedad de que las perturbaciones gravitatorias dan cuenta de casi todo el cambio de velocidad requerido. El más importante de esos mecanismos es probable que lleve consigo varias clases de interacciones resonantes capaces de aumentar la excentricidad orbital del pequeño cuerpo de modo que caiga bajo la influencia gravitatoria de Marte.

En relación con ello, debe atribuirse particular importancia al gran grupo de asteroides cerca del borde interior del cinturón de asteroides, entre 2,1 y 2,6 U.A. del Sol, que ocupan órbitas que cortan el plano de la de Júpiter bajo ángulos pequeños y son completamente estables en la escala de tiempo de la edad del sistema solar. James J. Williams, del Laboratorio de Propulsión a Chorro del Instituto de Tecnología de California, ha demostrado que las órbitas de esos asteroides tienen semiejes mayores cuyos valores son tales que su velocidad de precesión está casi en resonancia con la velocidad de precesión del sistema acoplado de los planetas mayores, dominados por Saturno y Júpiter. Bajo la influencia de sus vecinos gigantes, los asteroides de este grupo particular han sido conducidos a órbitas tales que, aunque son estables, varían considerablemente en excentricidad.

La misma estabilidad, sin embargo, no estaría garantizada para fragmentos del tamaño Apolo expulsados de tales asteroides por un choque accidental. Si la velocidad del fragmento aumentase en una cantidad tan pequeña como 200 metros por segundo y se situase en una órbita cuyo semieje mayor fuese sólo ligeramente más pequeño que el del cuerpo que lo originó, caería en la región crítica en que las resonancias con el sistema acoplado Júpiter-Saturno empezarían a aumentar la excentricidad orbital del fragmento, conduciéndolo con el tiempo a posiciones ocasionalmente muy próximas a Marte cuando este planeta estuviera cerca de su afelio.

Como resultado de tales pasos próximos, las perturbaciones gravitatorias inducidas por Marte causarán fluctuaciones aleatorias del semieje mayor del potencial objeto Apolo llevándolo a una posición donde la excentricidad orbital del objeto se amplificará por ulteriores interacciones resonantes hasta que finalmente corte la órbita de la Tierra y se convierta en un verdadero objeto Apolo. El proceso completo requeriría menos de 100 millones de años, y sin duda algunos Apolos se originan de esta manera. La mejor estima, sin embargo, es que tal proceso daría cuenta de algo así

como un solo objeto Apolo por cada millón de años. Habría que estirar mucho las probabilidades para justificar los 15 Apolos.

La otra hipótesis posible, que es la que yo defiendo basándome en todas las indicaciones disponibles, es que la mayor parte de los objetos Apolo proceden de antiguos cometas. Se sabe que los cometas emiten grandes volúmenes de material volátil, principalmente vapor de agua y dióxido de carbono. Presumiblemente, esas substancias están en su origen presentes en el núcleo del cometa, bien al estado sólido o bien en forma de compuestos más complejos. Tales substancias volátiles, sin embargo, no son los únicos componentes del núcleo. Cuando el cometa se acerca al Sol, se puede detectar mucho polvo de grano fino escapando junto con los gases. Lo típico es que se libere tanto polvo como gas. Cuando los fragmentos más grandes de polvo no volátil son capturados por la Tierra, producen los chubascos meteoríticos cometarios.

Algunos meteoros que son casi ciertamente de origen cometario tienen pesos de varias toneladas. Tan masivos cuerpos difícilmente pueden ser arrastrados junto con la habitual corriente de gases volatilizados; por tanto, deben proceder de sucesos más catastróficos, quizás asociados con la rotura del núcleo cometario, un acontecimiento que se observa ocasionalmente. En la ausencia de tal catástrofe, los objetos más grandes simplemente permanecerían con el cometa, enriqueciendo el núcleo en materia no volátil.

Los cometas que aparecen por primera vez han empleado gran parte de su vida viajando mucho más allá del más lejano planeta. Se hacen luminosos por primera vez cuando su órbita, perturbada por una estrella vecina, ha sido convertida en una con perihelio próximo al Sol. En la mayor parte de los casos, las perturbaciones ulteriores inducidas por los planetas serán la causa de que los cometas escapen del sistema solar mucho antes de que sus hielos volátiles se agoten. Una pequeña proporción de tales cometas de largo período, sin embargo, son desviados hacia órbitas estables más pequeñas que típicamente los conducen cerca del Sol cada siete años o así.

En el curso de cientos o miles de pasos por el perihelio, el núcleo del cometa va siendo gradualmente despojado de substancias volátiles. Se ha observado que tales cometas de corto período son marcadamente menos activos que los cometas que describen órbitas grandes y excéntricas. En unos pocos casos, los cometas de corto período han dado la impresión de ser enteramente "asteroidales" en el sentido de que durante algunos pasos por el perihelio no muestran coma o cola. No es por tanto una extrapolación muy grande imaginar que un cometa puede perder todas sus substancias volátiles y aparecer simplemente como un asteroide.

La desaparición de productos volátiles, sin embargo, no convierte por sí sola a un cometa en un objeto Apolo típico. Casi todos los cometas de período corto tienen afelios más allá de la órbita de Júpiter y todos los Apolos conocidos tienen sus afelios dentro de esa órbita. No obstante, hay excepciones. Varios cometas de corto período tienen afelios ligeramente interiores a la órbita de Júpiter. Uno de ellos, el cometa Encke, tiene su afelio muy al interior de ella, a 4,1 U.A. del Sol.

La actividad del cometa Encke ha disminuido visiblemente durante los tiempos históricos, y el cometa parece estar claramente en camino de convertirse en inactivo. Si lo hace completamente, quizás en el curso de unos pocos siglos, no será discernible de un objeto Apolo de bajo albedo. El Apolo más recientemente descubierto, el 1978 SB, es un objeto de tal clase y ocurre que su órbita es muy similar a la del cometa Encke. Muy posiblemente, el cometa Encke y el 1978 SB estén genéticamente relacionados. La existencia de chorros meteóricos en órbitas semejantes de corto período que dan lugar a meteoritos con propiedades físicas esencialmente idénticas a las de los meteoritos que se cree proceden del cometa Encke constituye un argumento a favor de que la evolución orbital del cometa no sea un hecho aislado. Recientemente, Whipple y su colega Zdenek Sekanina han presentado modelos físicos capaces de explicar tal evolución orbital. Demuestran que, en un cometa en giro, los cambios orbitales pueden producirse como consecuencia de la reacción dinámica al escape de los gases y el polvo.

La indicación suministrada por un cometa –el Encke– y varios chorros me-

INCLINACION PROPIA (GRADOS)
SEMIEJE MAYOR (UNIDADES ASTRONOMICAS)
REGION FLORA

FUENTE POTENCIAL DE OBJETOS APOLO, que es un grupo de cuerpos que están reunidos en órbitas de baja inclinación cerca de la parte más interna del cinturón principal de asteroides entre Marte y Júpiter. Tales asteroides se designan aquí por sus números de catálogo; los círculos llenos adyacentes indican su tamaño relativo. Se puede estimar que el grupo abarca unos 5000 cuerpos (*área sombreada*) que pertenecen a la región Flora (llamada así por el cuerpo más grande en esa región, el asteroide N.º 8). Aunque todos los objetos en el grupo ocupan órbitas estables, quedan cerca de la "superficie" de inestabilidad indicada por la curva negra. En 1969, James J. Williams, descubrió que los cuerpos cuya inclinación y semieje mayor los colocaba en esa superficie serían perturbados por Júpiter y lanzados a órbitas fuertemente excéntricas que cortan a la de la Tierra. Cuando los asteroides del grupo indicado aquí chocan, lo que debe suceder de vez en cuando, los fragmentos del tamaño Apolo podrían ser enviados a órbitas que caigan sobre, o en la proximidad, de la superficie de inestabilidad.

teóricos cometarios parece ofrecer solamente una base muy endeble para creer que los objetos Apolo sean de origen cometario.

En realidad, aun cuando todos los Apolos fueran cometas extintos, la existencia actual de un cometa activo tal como el cometa Encke es quizá más de lo que uno tiene derecho a esperar. Es probablemente un accidente afortunado el que el cometa Encke no desapareciese hace unos cuantos siglos, en cuyo caso no habría base observacional convincente para suponer que los cometas pueden convertirse en Apolos, aun cuando lo hagan así. Un único cometa como el Encke cada 65.000 años suministraría Apolos al ritmo de 15 cada millón de años, precisamente el necesario para mantener la población presente en régimen estacionario. Como la vida activa de un cometa de corto período es solamente de unos pocos miles de años, la mayor parte del tiempo no debería haber cometas activos en órbitas similares a la del Encke.

Mirando hacia el futuro, es posible que los objetos Apolo se puedan emplear para usos prácticos. Algunos de los objetos están en órbitas tan similares a la de la Tierra que resultan tan accesibles como la Luna. Además, son suficientemente pequeños para que se pueda considerar la posibilidad de usarlos como estaciones espaciales prefabricadas y maniobrables. En un esquema originalmente propuesto por Arthur C. Clarke y examinado en detalle por Henry H. Kolm, del Instituto de Tecnología de Massachusetts, los asteroides serían equipados con "impulsores de masa", motores que consumirían la propia substancia del asteroide para propulsarlo hacia nuevas órbitas. Otros han sugerido que se podría extraer de los Apolos la materia prima para la construcción de vehículos espaciales.

Antes de que tan exóticos proyectos sean puestos en marcha, puede ser conveniente pensar que la cantidad total de materia en todos los Apolos es solamente unos 10^{19} gramos, masa equivalente a no más que los dos o tres centímetros superiores de la superficie continental de la Tierra. Es probable que los Apolos, junto con unos pocos cometas, representen la única fuente no terrestre de compuestos de carbono y agua en el espacio vecino, fuente que requiere 25 millones de años para ser rellenada nuevamente. Quizá fuera mejor que la región del espacio ocupada por los objetos Apolo y sus parientes próximos, los objetos Amor, fuese declarada "reserva natural" inviolable.

Metástasis cancerosas

La metástasis, y no el tumor primario, constituye la gran amenaza del cáncer. Sólo determinadas células malignas poseen características que les permiten desplazarse por todo el cuerpo y originar nuevos tumores

Garth L. Nicolson

La mayoría de los enfermos cancerosos no mueren por el tumor primitivo, sino que fallecen por metástasis, es decir: por múltiples colonias de tumores que crecen en distintas partes del organismo, se originan a partir de células malignas que se separan del tumor inicial y se desplazan por el cuerpo, con frecuencia hasta lugares muy alejados. Si el tumor desencadenante se detecta precozmente, puede extirpársele por vía quirúrgica, radioterapéutica, quimioterapéutica o una combinación de tales tratamientos. Por desgracia, cuesta muchísimo más detectar y eliminar las colonias metastásicas. Desde un punto de vista clínico, podemos considerar la metástasis como el resultado final de la historia natural del cáncer. Además, la capacidad de metastatizar es propiedad característica, de modo exclusivo, de los tumores malignos.

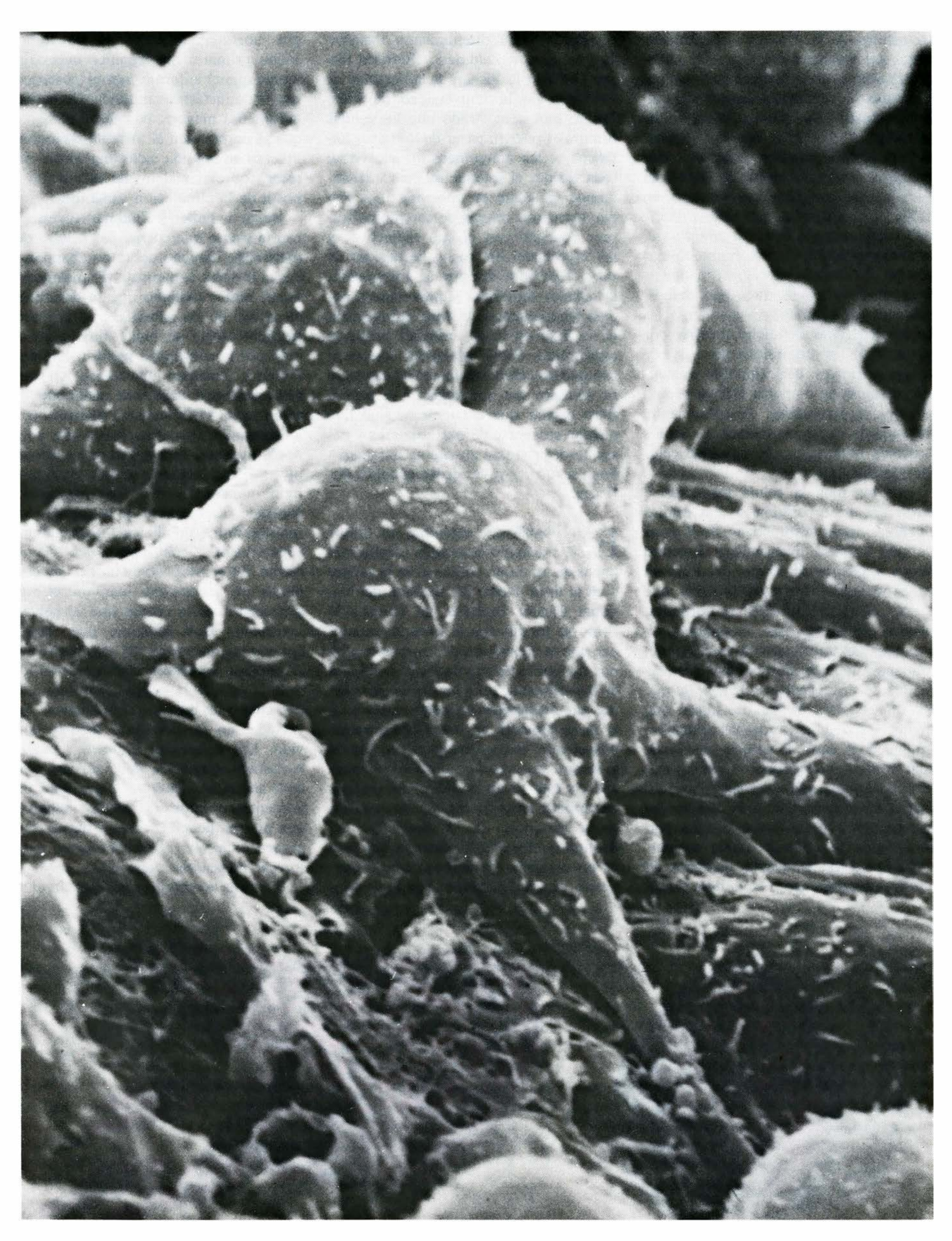
Surge el tumor de la transformación de una célula normal en otra que escapa a los controles usuales que tiene el hospedante sobre el crecimiento y la diferenciación celular. La célula transformada prolifera para formar un tumor, que puede ser benigno o maligno. Suele considerarse que un tumor es benigno cuando permanece con una estructura similar al tejido del que deriva, si crece lentamente por simple expansión y permanece encapsulado por una capa de tejido conjuntivo y, por último, si los núcleos de sus células se dividen casi normalmente, con pocos cromosomas anormales. Por contra, el tumor maligno es de estructura muy desigual de la del tejido originario, crece rápidamente y no permanece encapsulado, muestra muchas divisiones nucleares y cromosómicas anormales y, por último, invade el tejido normal circundante, diseminando células que tienen la propiedad de colonizar nuevos tejidos. Habida cuenta, pues, que los rasgos distintivos entre el estado benigno y el estado maligno que se basan en la velocidad de crecimiento y en el grado de "anormalidad" no son absolutos, diremos que lo que define específicamente a un tumor maligno, o cáncer, es la capacidad de invadir el tejido circundante y colonizar sitios alejados de su centro de formación. Por consiguiente, reviste capital importancia la comprensión de la biología de la metástasis para entender la naturaleza del cáncer y saber cómo prolongar o salvar las vidas de los enfermos cancerosos al destruir las células metastásicas, que son las más amenazadoras para la vida.

La metástasis cancerosa comprende una compleja serie de acontecimientos que implican interacciones entre las células malignas del tumor originario y células normales del organismo. Irving Zeidman, de la Facultad de Medicina de la Universidad de Pennsylvania, ha establecido las etapas secuenciales que concluyen en la colonización de un tumor metastásico distante: (1) extensión a los tejidos circundantes, (2) penetración en las cavidades y vasos del cuerpo, (3) desprendimiento de células tumorales para su desplazamiento a otras zonas, (4) reinvasión del tejido en el punto de parada y (5) alteración del nuevo tejido en orden a facilitar la supervivencia, vascularización y desarrollo tumoral de las células tumorales.

Como primer paso, las células malignas se extienden desde su localización original e invaden los tejidos normales inmediatos. Algunas células invasoras penetran en una cavidad del cuerpo (verbigracia, la cavidad abdominal o la torácica) o en el torrente sanguíneo, linfa o líquido espinal, independizándose del tumor primitivo. En los vasos sanguíneos o linfáticos, las células individuales o los grupos celulares (que se denominan émbolos tumorales) pueden verse transportados hasta lugares muy alejados de su origen. Las células que sobreviven al transporte en el sistema linfático suelen detenerse en los ganglios; las que sobreviven al torrente sanguíneo paran en las vénulas o capilares. Una vez asentadas, las células malignas pueden escapar del entorno hostil de la linfa o la sangre mediante la invasión del tejido adyacente. Por último, la nueva micrometástasis se vascularizará, y su nueva red de vasos sanguíneos aportará los nutrientes que le habrán de permitir la proliferación y la formación de una masa tumoral mayor. A lo largo del proceso metastásico, las células malignas deben resistir la destrucción por parte de los distintos mecanismos de defensa antitumoral del hospedante.

Se han avanzado varias teorías para explicar esta primera etapa de la metástasis: la invasión de los tejidos adyacentes. George Eaves, de la Facultad de Medicina de la Universidad de Leeds, sostiene que el comportamiento invasor se justifica, en sus líneas generales, por consideraciones mecánicas. Si inyectaba un plástico dental de fácil detención en muestras tisulares, el plástico infiltraba las zonas de menor resistencia, un comportamiento parecido al que manifiestan la mayoría de los tumores malignos. Pero no parece que los simples modelos mecánicos puedan descifrar el comportamiento invasor, extremadamente rápido, de algunos tumores; los experimentos con cultivos de tejidos demuestran que las células malignas de gran poder invasor llegarán a penetrar una lámina de células normales incluso en ausencia de cualquier presión mecánica. Dale R. Coman, de la Facultad de Medicina de la Universidad de Pennsylvania, opina que la capacidad invasora en los tejidos circundantes les viene a las células malignas de su movilidad; pero también es cierto que hay células normales con gran motilidad y carecen de dicha capacidad invasora.

La presencia en el tejido tumoral de elevadas concentraciones de ciertos enzimas degradadores dio pie a que se pensara que la conducta invasora de las células tumorales pudiera facilitarse con la destrucción enzimática de la matriz de tejido conjuntivo normal que media en-



CELULAS METASTASICAS DEL MELANOMA invadiendo la pared de un vaso sanguíneo. Las células del melanoma del ratón, que metastizan, típicamente, por vía sanguínea, se incubaron con un fragmento de tejido venoso y se mantuvieron en él en un cultivo de laboratorio. Las células melanómicas se han adherido a las epiteliales (éstas revisten la superficie interna de la vena) y han introducido sus procesos en las fisuras que han abierto a raíz de que las células epiteliales se replegaran ante el ataque de las tumorales. La adhesión al endotelio y la invasión del mismo constituyen dos fases del proceso por el que una célula tumoral introducida en el torrente sanguíneo establece una nueva colonia tumoral. Aumento: x 7000.

tre células. A. R. Poole y sus colaboradores del Strangeways Research Laboratory compararon enzimas que digieren proteínas y que son segregados por cultivos celulares de tumores benignos de la mama de la mujer con los enzimas segregados por tumores malignos y por tejidos normales de mama. Bengt Sylvén llevó a cabo experimentos similares en el Karolinska Institute, en Suecia. Poole halló que una proteinasa, la catepsina B, se acumulaba en proporción mucho más alta en los cultivos de células malignas. Pero otras investigaciones han demostrado la existencia de escasas diferencias en la liberación de dicho enzima por parte de tejidos tumorales y por parte de tejidos normales, de forma que la invasión en cuestión no podía explicarse sólo por acción enzimática. Probablemente, depende de varias combinaciones de los procesos mecánicos, de motilidad celular y enzimáticos.

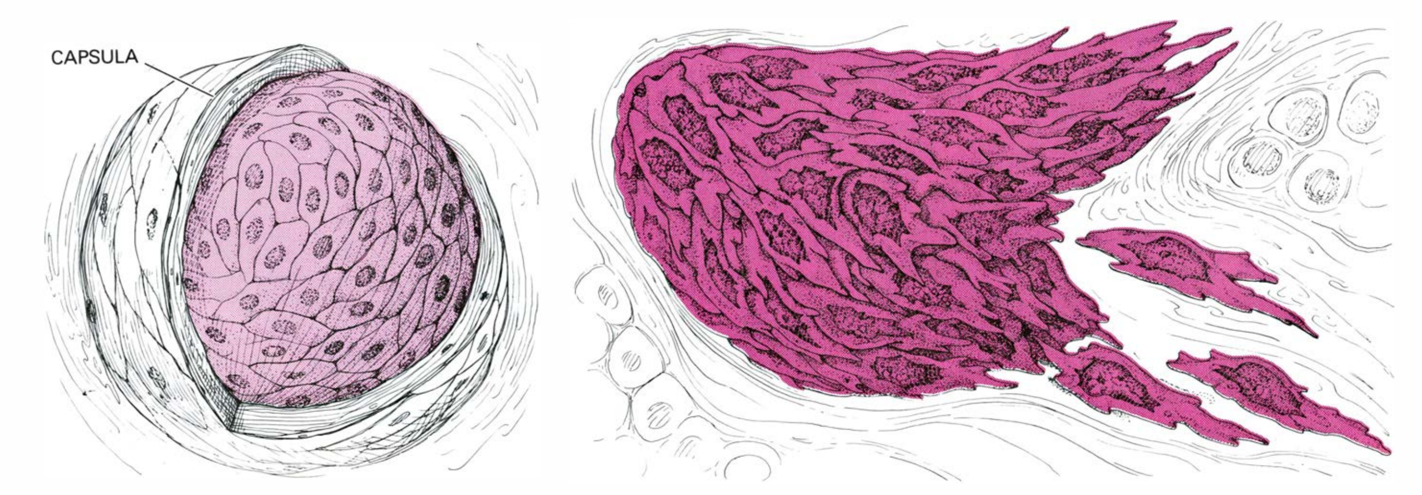
Ya tenemos la célula extendida a partir del tumor originario. Las células malignas deben desprenderse de él antes de que puedan metastatizar. Leonard Weiss y George H. Poste, del Roswell Park Memorial Institute, afirman que ciertas células, tumorales o normales, pueden destruirse, lisar, incrementando los niveles extracelulares de enzimas degradantes que liberan racimos de células malignas. Hay pruebas, además, de que las fuerzas de cohesión normales que tienden a mantener unidas las células remiten en el caso de células malignas. Así, Coman observó que si agitaba racimos de células malignas del cáncer de mama se desprendían más celulas del grupo que si la agitación ocurría con fragmentos de tejido de mama normal.

La liberación de células malignas en el torrente sanguíneo o en la linfa puede resultar en una diseminación metastásica del tumor. Zeidman y Bernard Fisher y Edwin R. Fisher, de la Facultad de Medicina de la Universidad de Pittsburgh, han demostrado que las células tumorales pueden pasar de los vasos linfáticos a los sanguíneos, y viceversa, toda vez que el sistema linfático y el sanguíneo se hallan en conexión. Los Fisher marcaron células de carcinoma (tumor epitelial) con cromo radiactivo, y las inyectaron en venas periféricas de ratas. Una hora después de la inyección, las células tumorales supervivientes podían recuperarse de la linfa; la inyección de la linfa en otros animales producía carcinomas. Estos resultados demostraron que las células malignas de la sangre no siempre son atrapadas en el primer filtro capilar que encuentran, sino que pueden escaparse de esos capilares y pasar desde la sangre al sistema linfático, a buen seguro, a través de los ganglios linfáticos.

Es probable que mientras un tumor esté creciendo desprenda células malignas de un modo continuo en la sangre o en la linfa. Pietro M. Gullino, del National Cancer Institute, ha controlado cómo se modificaba el torrente sanguíneo con la deposición de tumores malignos de mama trasplantados en animales de experimentación y ha observado que, tras un crecimiento en pocos gramos de la masa tumoral, el tumor libera varios millones de células malignas por día en la sangre. La presencia de células malignas en ésta (o en la linfa) no siempre desemboca en la formación de metástasis a distancia. A. J. Salsbury, del Brompton Hospital londinense, fue uno de los primeros en destacar que la presencia de células cancerosas en la sangre, liberadas a partir de un tumor sólido primario, no significa necesariamente que el enfermo morirá posteriormente de metástasis múltiple. De hecho, la inmensa mayoría de células malignas que se desprenden al torrente sanguíneo mueren en seguida. Sólo unas pocas parecen sobrevivir al medio turbulento y hostil de la circulación sanguínea.

Isaiah J. Fidler y Zeidman investigaron la suerte que corrían las células del melanoma maligno del ratón abandonadas al torrente sanguíneo. (El melanoma es un tumor maligno que se desarrolla en la piel de los animales y el hombre; frecuentemente se halla muy pigmentado por la substancia negra melanina.) Fidler y Zeidman inyectaron, en una vena periférica, células de melanoma de ratón cuyo ADN se había marcado con ioduro radiactivo. Observaron que, a los pocos minutos, la mayoría de las células inyectadas se habían detenido en el pulmón, que era el primer filtro capilar encontrado. La mayoría de las células tumorales murieron allí; sólo unas cuantas supervivientes continuaron en el torrente. En el intervalo de un día, sin embargo, no sobrevivían más que el uno por ciento de las células que se habían inyectado. Dos semanas más tarde, cuando las metástasis del melanoma negro podían apreciarse claramente en los pulmones, sólo la décima parte del uno por ciento de las células radiactivas inyectadas en un comienzo seguían con vida. Los resultados indicaron que la metástasis sanguínea debe ser un proceso altamente selectivo y que sólo un porcentaje mínimo de células malignas que pueden penetrar en el torrente sanguíneo y circular en él sobreviven para formar metástasis a distancia.

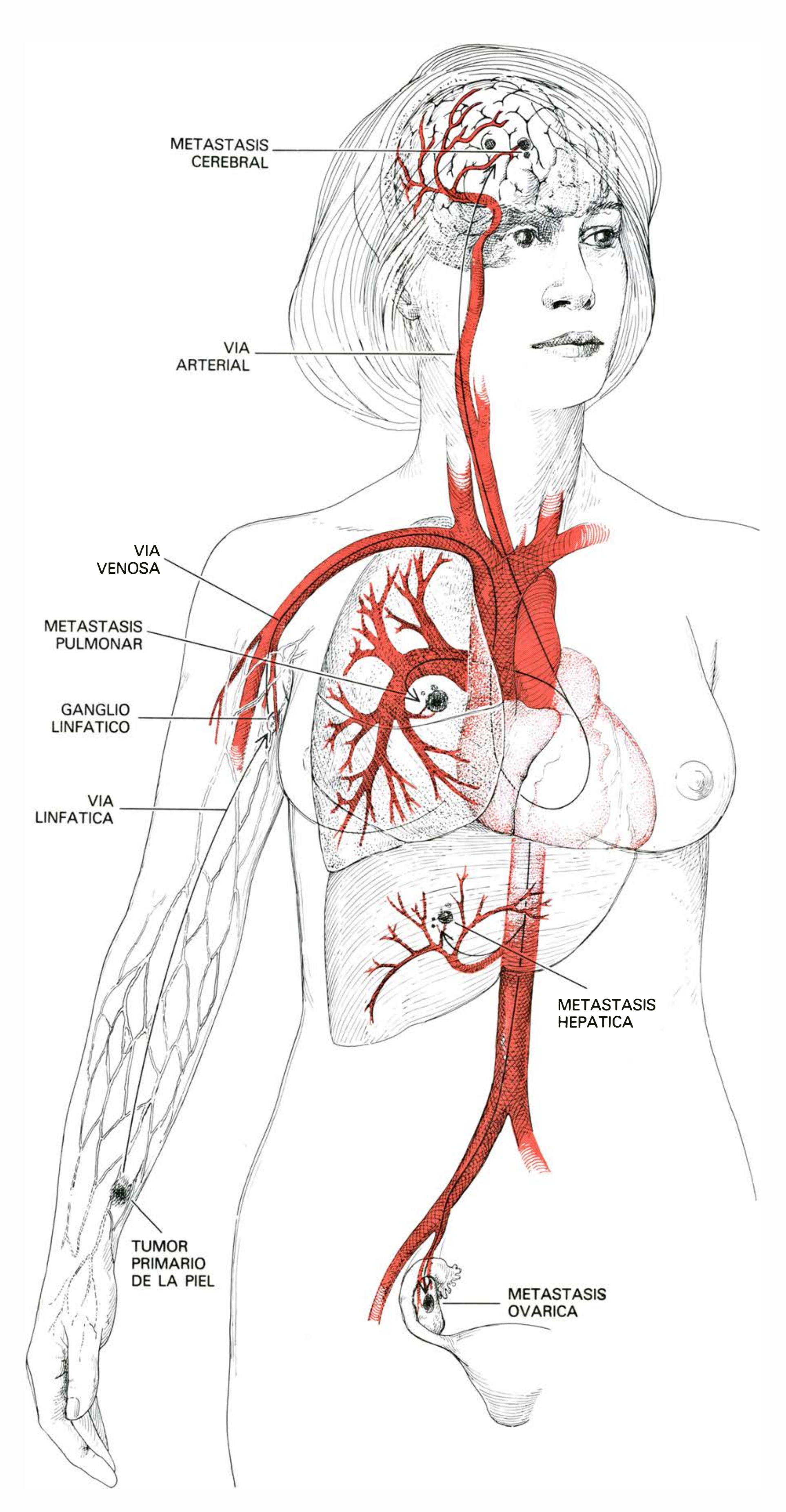
Mientras las células malignas circulan



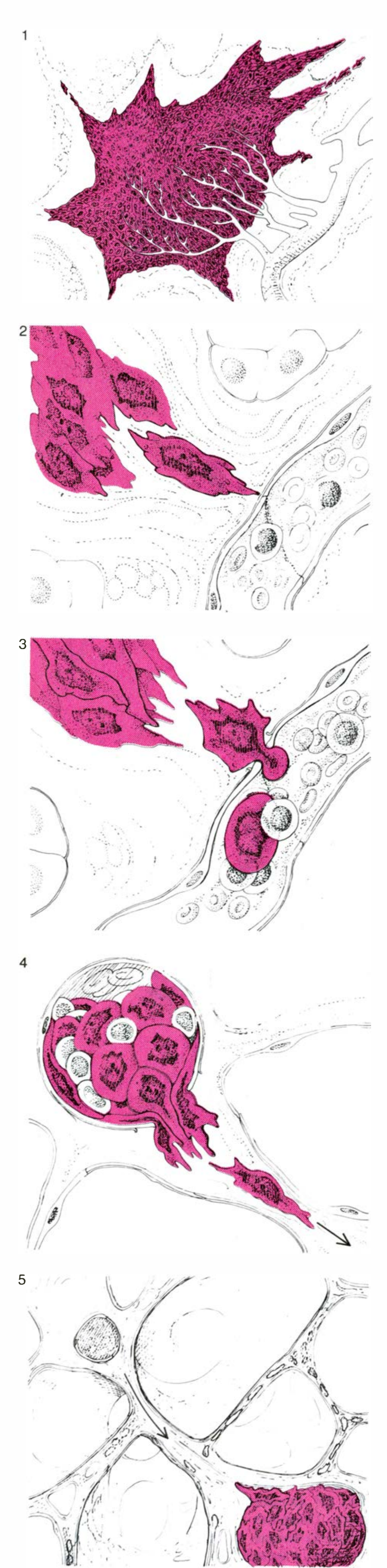
SE DEFINE EL TUMOR como una masa de tejido que surge de la proliferación de una célula que ha sufrido tal transformación que logra escapar de los controles normales de desarrollo y diferenciación. En general, un tumor benigno (*izquierda*) aparece casi normal en su estructura hística, crece poco a poco y permanece encapsulado por tejido conjuntivo. No invade el tejido circundante, ni sus células se desprenden para metastatizar. Un tumor maligno (*derecha*) posee una estructura atípica, crece deprisa, no se halla encapsulado e invade el tejido circundante; algunas células se desprenden.

por la sangre, llevan a cabo múltiples e importantes interacciones celulares, según se ha demostrado en mi laboratorio de la Universidad de California en Irvine y en los laboratorios de Fidler (quien se encuentra hoy en el National Cancer Institute's Frederick Cancer Research Center), de Gabriel J. Gasic y Tatiana Gasic, en la Facultad de Medicina de la Universidad de Pennsylvania, y de Lance A. Liotta, en el National Cancer Institute. Dichas interacciones inciden en la supervivencia de las células y en su detención en los capilares. Las células malignas del torrente sanguíneo pueden, por ejemplo, unirse a otras células tumorales para constituir racimos o grupos. Pueden adherirse también a células sanguíneas normales, como los linfocitos y las plaquetas. Y cuando penetran en un lecho capilar, pueden unirse a células endoteliales, que revisten el interior de los vasos. Cuanto mayor es el grupo, más probable se hace su detención a nivel de los capilares, acabando por formar una nueva colonia tumoral. Una vez retenidas en los capilares, algunas células tumorales estimulan la producción de fibrina, que es la proteína fibrosa que está implicada en la formación de coágulos sanguíneos. B. A. Warren, de la Universidad americana de Western Ontario, ha observado que la trama resultante de fibrina tiende a proteger las células malignas, permitiéndoles la proliferación.

Las células malignas se han detenido en las vénulas o en los capilares. Ahora, tendrán que atravesar la pared del vaso sanguíneo y penetrar en su tejido circundante para sobrevivir. La fuga de las células tumorales de la circulación ha sido investigada por Randall Kramer en mi laboratorio. Según parece, las células malignas inducen un cambio drástico en la cohesión de las células endoteliales (las cuales revisten la superficie interior de los vasos sanguíneos) con su matriz extracelular subyacente; es decir: las células endoteliales se retraen, dejando espacios libres entre sí, por donde las células tumorales pueden escapar del sistema vascular. Una vez las células tumorales han invadido el tejido extravascular, puede implantarse un nuevo microentorno. El desarrollo de la neocolonia tumoral se frena hasta que puede disponerse de un nuevo aporte sanguíneo que acarree los nutrientes y depure los productos de desecho. La vascularización del nuevo tumor se ve estimulada por "factores angioformativos tumorales" liberados por las células malignas. Se supone que tales factores actúan mediante la estimulación especí-



DISEMINACION METASTASICA de un melanoma maligno (cáncer de la piel pigmentado). Puede desencadenarse cuando células malignas de un tumor, del antebrazo por ejemplo, invaden un vaso linfático y la linfa los acarrea hasta un ganglio linfático. Detenidas transitoriamente allí, las células pueden penetrar en un vaso sanguíneo; luego, a través de la sangre, alcanzan el corazón y de éste a otras partes del cuerpo. Las células del melanoma propenden a metastatizar el cerebro, pulmones, etc.



fica de la proliferación de células endoteliales por división celular y su desplazamiento para formar nuevos capilares.

El proceso metastásico depende, luego, de varias y afortunadas interacciones entre las células malignas y las células normales del hospedante. Es probable que muy pocas de las células malignas del tumor desencadenante puedan llevar a cabo estas interacciones y cubrir, por tanto, la secuencia de etapas de la metástasis. Sólo una pequeña fracción de las células que se desprenden del tumor originario invaden una cavidad corporal o penetran en el sistema linfático o sanguíneo, y se independizan. De estas células sólo unas cuantas sobreviven al transporte. Menos todavía serán las que pervivirán hasta completar el proceso metastásico, ya que en cada fase se hallarán expuestas a los mecanismos de defensa del organismo.

El que sólo una fracción pequeña de células tumorales alcancen a formar nuevas colonias metastásicas da a entender que aquellas que lo logran deben tener unas cualidades especiales. ¿Acaso representan una subpoblación de células tumorales con características determinadas que les capacita para afrontar nuevas metástasis con éxito? ¿Se debe, tal vez, a la firme presión de las defensas del hospedante esa selección natural de dicha subpoblación de células malignas con propiedades altamente metastatizantes?

Esta hipótesis puede someterse a comprobación en modelos animales con cáncer metastásico. Una de las formas más útiles es el melanoma de ratón. El melanoma metastásico constituye uno de los cánceres menos susceptibles de tratamiento en el hombre, debido en gran parte a la frecuencia de metástasis linfáticas y sanguíneas; e igual ocurre con los melanomas no humanos. El melanoma *B*16, que apareció en una cepa de ratones negros seleccionada en el Jackson Laboratory, puede trasplantarse de un ratón a otro y, también, puede desarrollarse en cultivo de tejidos. Cuando las células del melanoma *B*16 se implantan en la piel de ratón, lo normal es que el melanoma metastatice a una tasa baja o moderada. Para afinar más el experimento sobre la capacidad de metástasis, suele inyectarse cantidades concretas de células tumorales de modo directo en el sistema circulatorio, en la vena de la cola. Las células circulan, se detienen, invaden, sobreviven y proliferan hasta formar varias colonias tumorales pigmentadas en los pulmones, y, en menor número, en los ovarios, ciertas cavidades del organismo, hígado y cerebro.

Fidler, mientras se hallaba en la Escuela de Odontología de la Universidad de Pennsylvania, llevó a cabo un experimento en el que se sirvió de células del melanoma *B*16 para seleccionar subpoblaciones con capacidad potenciada de colonizar los pulmones. Comenzó con la extirpación quirúrgica de las colonias de melanoma que formó en los pulmones de los ratones en que se habían inyectado, por vía intravenosa, células *B*16. Hizo crecer estas células tumorales en un medio de cultivo tisular; sacó luego las células cultivadas, a las que se designó como línea celular variante del melanoma *B*16-*F*1; y las inyectó luego por vía intravenosa en otro grupo de ratones. Las nuevas colonias de tumores pulmonares experimentales que resultaron de esta segunda selección se volvieron a cultivar para obtener un nueva línea celular, la *B*16-*F*2. El proceso de inyección, colonización pulmonar, cultivo del tejido y recogida para su reinyección en ratones prosiguió. En cada ciclo subsiguiente, aumentó la eficiencia de detención, por parte de las distintas líneas, en los pulmones, su posterior supervivencia, invasión y desarrollo hasta formar colonias tumorales pulmonares.

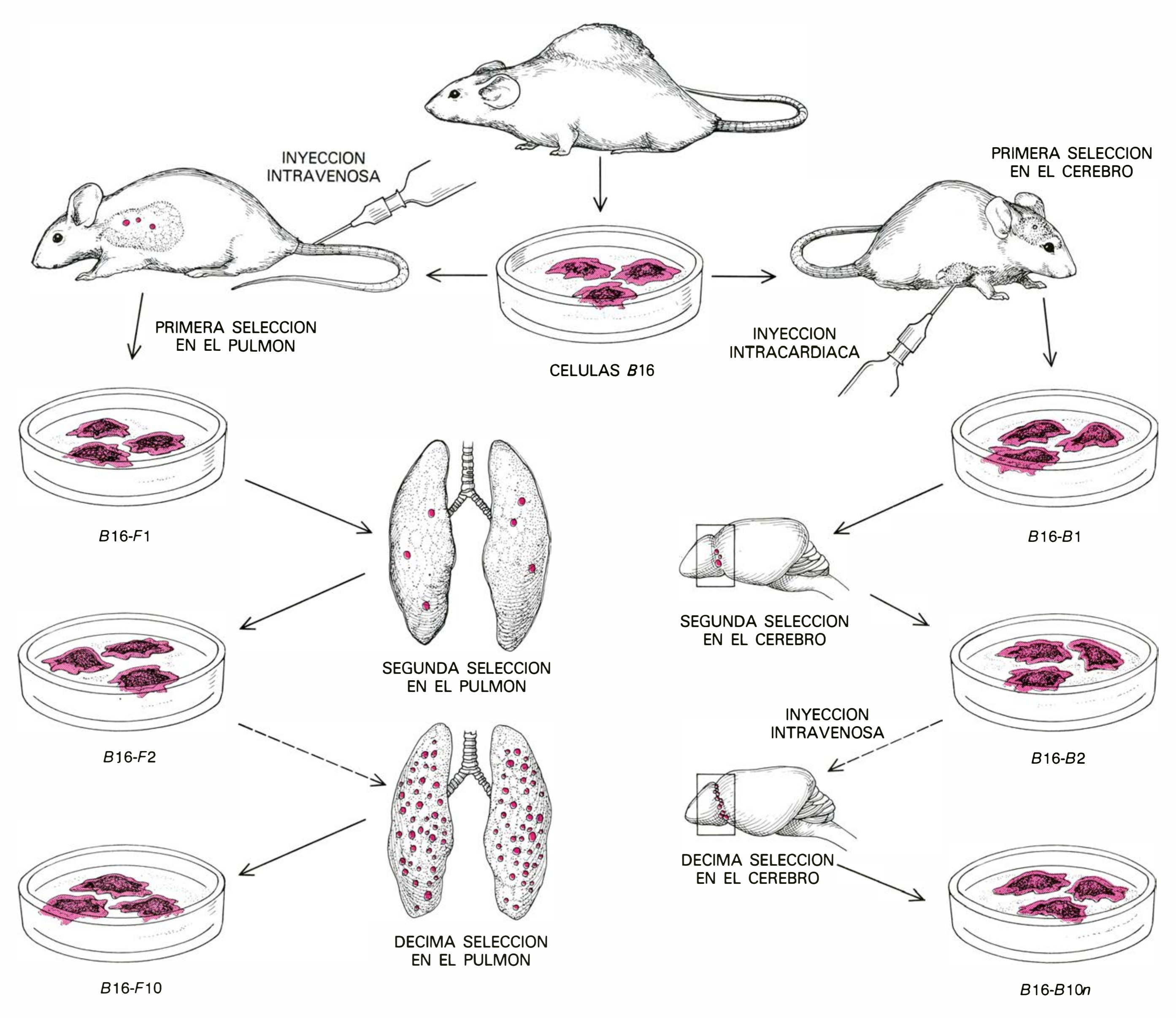
A los diez ciclos, Fidler estableció una línea melanómica variante, la *B*16-*F*10, que formaba un número significativamente mayor de colonias tumorales pulmonares que la línea parental *B*16 o que la primera línea seleccionada, la *B*16-*F*1, cuando se les inyectaba a los ratones un número dado de células malignas. Esta familia de líneas variantes de células tumorales que poseen diferentes potenciales metastásicos ha venido a constituir un eficaz instrumento en gran número de investigaciones biológicas y bioquímicas sobre las secuencias de metástasis tumorales. Su éxito como modelo para explicar la metástasis se debe, probablemente, a que la línea tumoral original *B*16 no constaba de una población celular uniforme, sino que, a buen seguro, era heterogénea con respecto a aquellas particulares características celu-

ETAPAS, esquematizadas, de una metástasis por vía sanguínea. El tumor primario aumenta a medida que proliferan las células malignas, pero sólo crece con rapidez cuando la vascularización aporta la cantidad necesaria de sangre (*1*). El tumor invade el tejido circundante (*2*). Algunas células, aisladas o en grupos, penetran en la membrana inferior de un capilar, discurren por entre las células endoteliales y se introducen en el torrente sanguíneo, en donde pueden interaccionar con células sanguíneas, los linfocitos por ejemplo (*3*). A través del corazón, las células tumorales llegan al pulmón, en donde se detienen en un capilar; las células tumorales atraviesan la pared capilar (*4*), entran en el tejido alveolar y proliferan hasta establecer una colonia tumoral (*5*).

lares tumorales que son necesarias para que ocurra la metástasis.

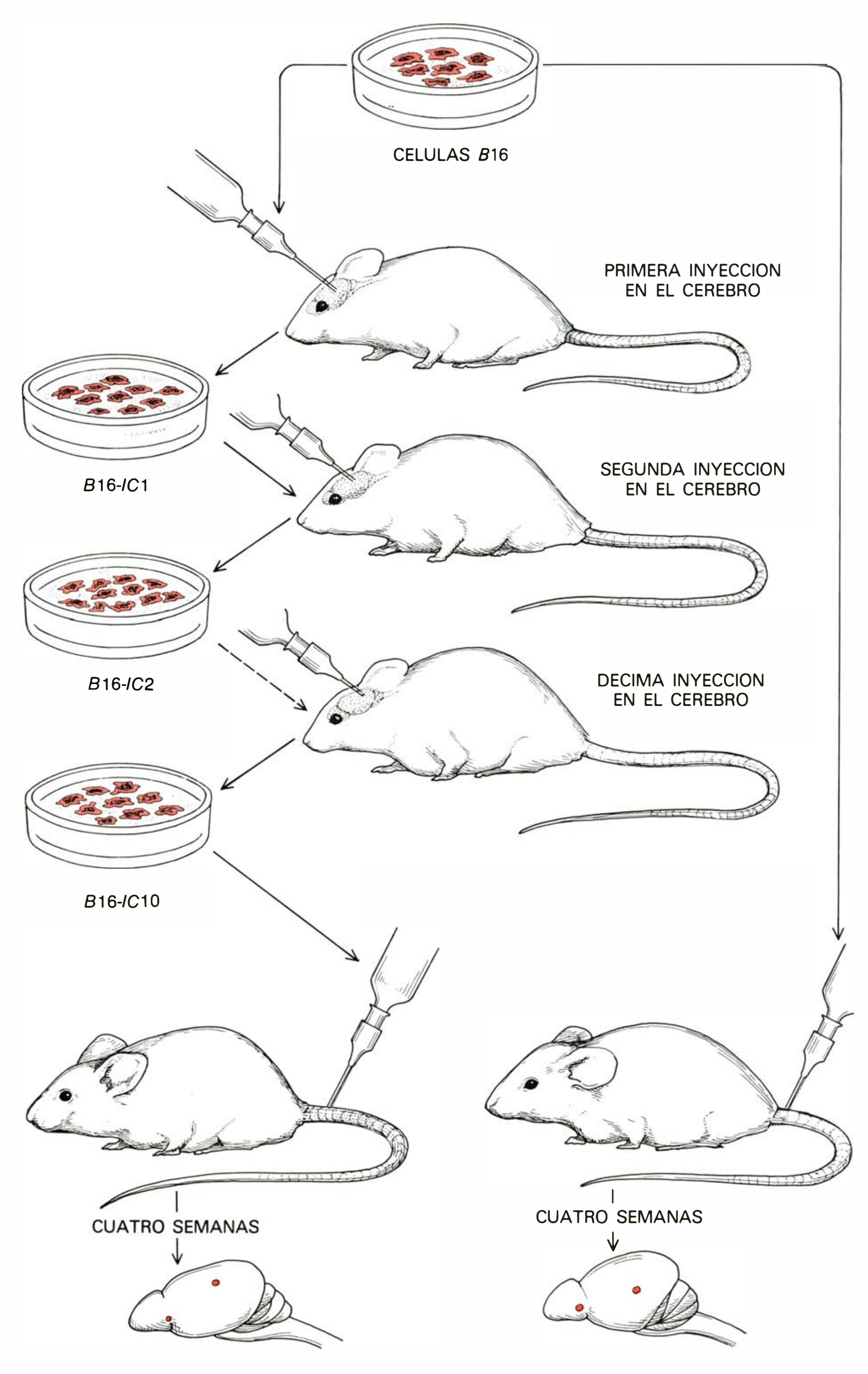
Ciertas observaciones clínicas denotan que determinados tipos de tumores primarios encierran una tendencia a metastatizar, de un modo preferente, en algunos órganos. Así, los carcinomas de mama tienden a diseminarse por el cerebro y pulmones; los tumores pulmonares desarrollan frecuentemente metástasis en el cerebro y glándulas suprarrenales; y el carcicoma de próstata lo hace en el hueso. También se han observado determinados modelos selectivos de metástasis para algunos tumores en animales. Si el modelo de distribución metastásica para un tipo celular dado no es aleatorio, ¿a qué se debe?

A principios de siglo, los cancerólogos James Ewing y Stephen Paget avanzaron ya una hipótesis para explicar la no-aleatoriedad de la metástasis. De acuerdo con la cual, la distribución metastásica venía determinada por condiciones anatómicas y vasculares así como por el microentorno para la supervivencia y desarrollo del tumor en el punto de implantación de la metástasis. Ello implicaba que las células tumorales proliferan en un "suelo" determinado si son capaces de adaptarse al mismo; esa es la razón por la que a la hipótesis se la ha denominado "hipótesis del suelo". Más recientemente, se ha puesto en claro que las propiedades de las células malignas influyen sobremanera en sus tipos de metástasis. Tal como Everett D. Sugarbaker y Alfred S. Ketchem, de la Facultad de Medicina de la Universidad de Miami, pusieron de relieve, la distribución y supervivencia de las distintas células malignas depende, probablemente, de un gran número de factores, a saber: las propiedades específicas de las propias células, el medio ambiente para su supervivencia y desarrollo en órganos concretos, así como el estatuto inmunitario antitumoral del hospedante.



SELECCION DE LINEAS de células melanómicas que poseen una incrementada capacidad de metastizar en los pulmones (*izquierda*), realizada por Isaiah J. Fidler, a la sazón en la Escuela de Odontología de la Universidad de Pennsylvania. Kenneth W. Brunson, del laboratorio del autor, seleccionó líneas que metastizaban en una región particular del cerebro (*derecha*). Para la selección pulmonar, Fidler tomó células melanómicas *B*16 de un cultivo de un tumor de ratón; las inyectó en la vena de la cola de ratones y recogió células de las colonias tumorales pulmonares resultantes. Estas células *B*16-*F*1 fueron cultivadas, a su vez, y vueltas a inyectar por vía intravenosa. Las células altamente seleccionadas *B*16-*F*10, producidas tras diez ciclos de éstos, formaron un número significativamente mayor de metástasis pulmonares que las no seleccionadas de la línea *B*16, y sólo llegaron a colonizar los pulmones. Por su parte, Brunson inyectó células *B*16 en el ventrículo izquierdo del corazón y extrajo los raros tumores cerebrales que se habían formado. Las células tumorales, denotadas *B*16-*B*1, se cultivaron y volvieron a inyectar de nuevo en el corazón; esta vez, los tumores fueron extraídos sólo del rinencéfalo, en las proximidades de la parte anterior del cerebro (*rectángulo*.) El proceso se repitió con las células inyectadas por vía intravenosa. Después de diez selecciones, Brunson obtuvo la línea *B*16-*B*10*n*, que es altamente metastizante sólo para el rinencéfalo.

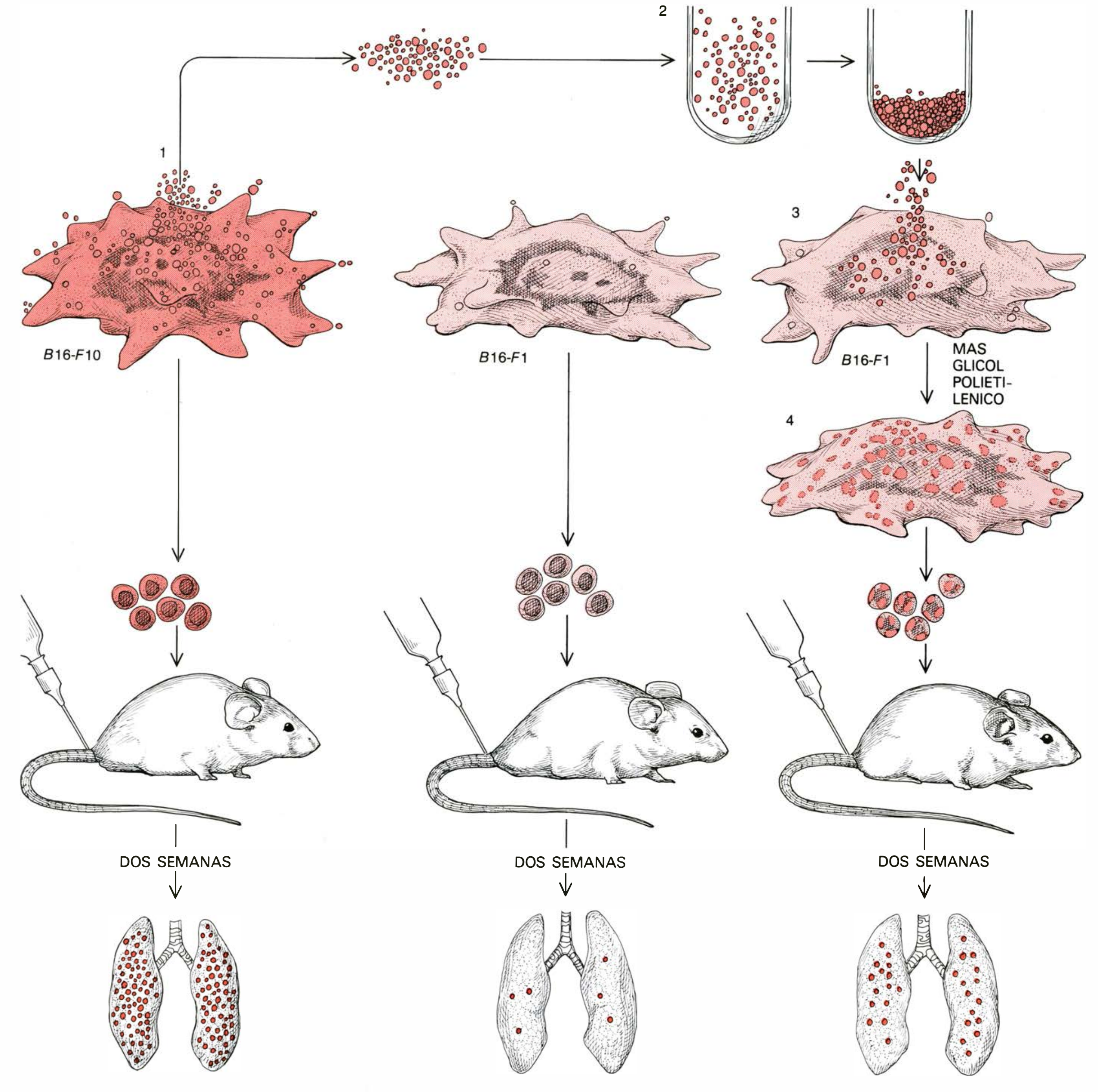
En colaboración con Fidler hemos estudiado la especificidad con que subpoblaciones de células melanómicas *B*16 de ratón se asientan en el pulmón y forman allí colonias. ¿Quedan las células tumorales atrapadas en el pulmón simplemente porque se trata del primer filtro capilar que encuentran? ¿O acaso buscan específicamente el tejido pulmonar? Inyectamos células *B*16-*F*1, de relativa especificidad, y células *B*16-*F*10, altamente específicas, en la vena de la cola (desde donde encontrarán primero los capilares pulmonares) y en el ventrículo izquierdo del corazón (que recibe sangre de los pulmones; en cuyo caso, las células inyectadas encontrarán otros lechos capilares antes de alcanzar los pulmones). La distribución inicial de las células tumorales radiactivamente marcadas fue tajantemente distinta, lo cual daba a entender que se habían detenido en un comienzo en los capilares primeramente encontrados a partir de su punto de inyección. Pero al día siguiente de la inyección, la distribución resultaba ya muy parecida, y el número de células pertenecientes a cada línea que seguían en vida, el mismo, y ello a pesar de ser diferente el punto de entrada en el torrente circulatorio. Al filo de las dos semanas, se registró casi la misma cantidad de tumores pulmonares para ambas rutas, con la salvedad que era de esperar: había un número significativamente mayor de tumores formados por células *F*10 que por células *F*1. Los resultados indicaban que las células tumorales destinadas a fundar metástasis en un órgano particular podían incluso despegarse del punto de detención inicial y volver a circular hasta que dieran con el entorno adecuado para fijarse, sobrevivir y desarrollarse.



COLONIZACION PREFERENCIAL DEL CEREBRO, que no depende de la adaptación al entorno cerebral. Esto se demostró en un experimento en que se inyectaron células *B*16 directamente en el cerebro, y no en el torrente sanguíneo. Los melanomas formados en el cerebro se extrajeron y se cultivaron sus células para volverlas a inyectar en el cerebro. Después de diez ciclos de éstos, las células recogidas, inyectadas en los ratones, no formaron más metástasis cerebrales que produjeron las células *B*16 originales cuando se las inyectó por vía intravenosa. Las células altamente metastásicas en el cerebro son una subpoblación de la línea *B*16 con propiedades que les permiten su detención allí.

La selección de especificidad por colonizar un determinado órgano no sólo parece significar que las células serán más eficaces en ese órgano particular, sino que, además, devienen menos aptas para colonizar otros órganos. En este experimento, así como en otros muchos, observamos que mientras las células *B*16-*F*1 formaban algunos tumores fuera del pulmón de ratón (principalmente en ganglios linfáticos, ovarios, cavidad torácica, hígado, intestinos, cerebro y glándulas suprarrenales), las células altamente seleccionadas *F*10 no lo hicieron. Sólo colonizaron el pulmón.

En mi laboratorio, Kenneth W. Brunson avanzó el proceso de selección un peldaño más. Comenzó con la línea *B*16-*F*1, que coloniza preferentemente el pulmón, aunque no de un modo exclusivo; a partir de ella seleccionó subpoblaciones cuya preferencia de colonización se inclinaba por puntos muy diferentes. En uno de sus experimentos extirpó tumores cerebrales, que constituye una rareza verlos formados por células *B*16-*F*1, adaptó las células a un cultivo tisular y obtuvo una línea que bautizó *B*16-*B*1. Prosiguió luego con el sistema de inyección, extracción del tumor y cultivo que Fidler había empleado para obtener la línea *F*10 específica del pulmón, con la variante de que, en cada etapa, extirpaba tumores cerebrales. Tras diez selecciones, Brunson había desarrollado una línea, la *B*16-*B*10*n*, que metastatiza una zona concreta del cerebro de ratón, el rinencéfalo, entre la cor-



SE VIO QUE LA MEMBRANA CEREBRAL desempeñaba un papel importante a la hora de determinar la capacidad metastásica de las células *B*16. Fragmentos de membrana que han sido espontáneamente liberados como vesículas (*1*) a partir de la membrana externa de células de la línea *B*16-*F*10, altamente metastásica (*color obscuro*), se recogieron de un medio de cultivo, se concentraron (*2*) e incorporaron en un cultivo de células *B*16-*F*1, mucho menos metastásicas (*3*). En presencia de glicol polietilénico, las vesículas *F*10 se fusionaron con las membranas de las células *F*1 (*4*.) Se inyectó a ratones las tres líneas celulares. Las células *F*1, cuyas membranas celulares contenían fragmentos de membrana de células *F*10, producían un número de tumores pulmonares significativamente mayor que el producido por las células *F*10 no tratadas, aunque no tantos como las células *F*10.

teza cerebral y el bulbo olfatorio. A través de un proceso similar, se seleccionó otra línea celular tumoral (*B*16-*O*10), que coloniza preferentemente los ovarios, un lugar habitual de metástasis melanómica humana.

La preferencia que muestran las líneas *B*16 por colonizar órganos como el cerebro, ¿resulta de la adaptación aleatoria de células melanómicas a un medio particular (como diría la hipótesis del suelo) o bien depende de propiedades específicas que poseen desde un comienzo ciertas células tumorales? Brunson hizo experimentos similares en los que células *B*16 y líneas celulares tumorales subsiguientes seleccionadas se implantaban directamente en el cerebro en lugar de inyectarlas en el torrente sanguíneo; con lo cual se eliminaban las fases de circulación sanguínea y detención en los capilares. Tras diez selecciones para la adaptación y crecimiento en el cerebro, las células melanómicas *B*16 recuperadas de los tumores cerebrales no se mostraban más eficaces en la metástasis del cerebro (cuando se las inyectó en la piel o en la circulación sanguínea) que la línea *F*1 inicial. En otras palabras, no se trata de una mera adaptabilidad al desarrollo en el interior del cerebro lo que se selecciona en la secuencia de *B*1 a *B*10*n*, sino la capacidad para asentarse en el cerebro, detenerse allí y colonizarlo.

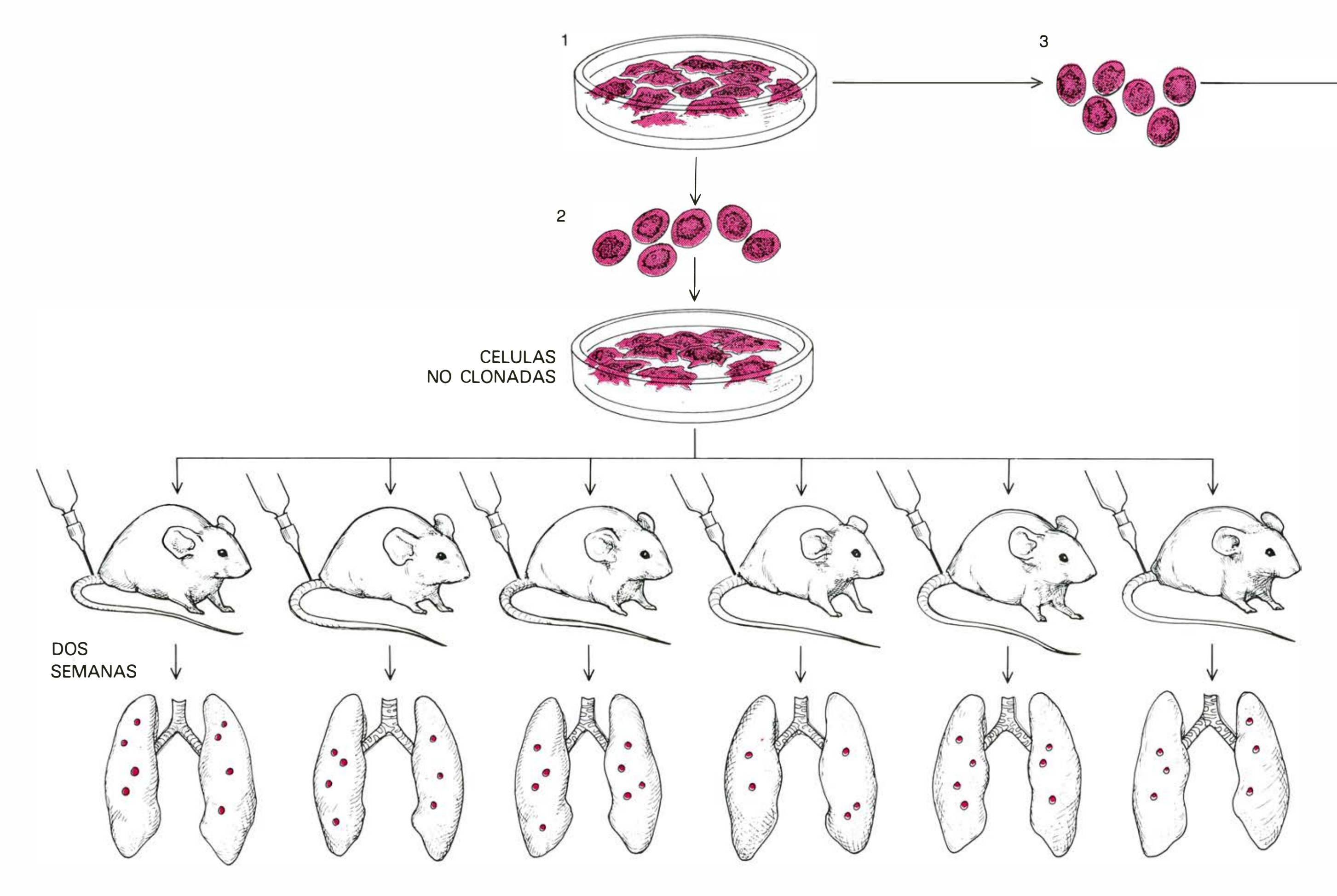
El éxito buscado en la identificación de las propiedades específicas de las líneas celulares tumorales que les permiten metastatizar lugares determinados podría suponer una notable primera

etapa en el desarrollo de un proceso de bloqueo de las metástasis en los pacientes cancerosos. Aunque estamos lejos de conocer qué propiedades celulares contribuyen a la eficacia de las metástasis, parece que la superficie de la célula tumoral se halla directamente involucrada en ciertos pasos, verbigracia, en la detención de la célula en un lugar determinado.

George Poste y el autor llevaron a cabo un experimento para demostrar la incidencia de la superficie celular, una vez que habíamos observado que las células *B*16-*F*10 de un cultivo tisular desprendían fragmentos de membranas celulares en forma de pequeñas vesículas cerradas. Recogimos las vesículas. Las purificamos por centrifugación en tubos que contenían un gradiente de una densa solución azucarada. Las vesículas purificadas se incorporaron a un cultivo de células *B*16-*F*1 con el polímero glicol polietilénico. En presencia del polímero, las vesículas *F*10 se fundieron con, y pasaron a formar parte de, la membrana celular de las *F*1. Cuando las células *F*1 enriquecidas con vesículas de membranas de *F*10 se inyectaron en ratones, metastatizaron de un modo más afín a como lo hace la línea, altamente metastatizante, *B*16-*F*10 que al modelo seguido por las células *B*16-*F*1 solas; ello ponía de manifiesto que la membrana celular se halla implicada necesariamente en la metástasis.

Entonces, ¿cómo se diferencian las membranas celulares en las líneas *B*16 que colonizan sitios distintos? Comparamos los componentes proteicos de las superficies de varias líneas celulares mediante el método de incorporar moléculas radiactivas en las proteínas, disolviendo las membranas celulares en soluciones de detergente y sometiendo a electroforesis las proteínas solubles de las membranas; las separábamos pues de acuerdo con su tamaño y su carga eléctrica. Cada una de la tres líneas melanómicas –la que coloniza el pulmón, la del cerebro y la de los ovarios– mostraron una configuración diferente de proteínas superficiales específicas; y, en general, la cantidad (o la accesibilidad) de dichas proteínas de superficie específicas aumentaba con la potencia metastásica de la célula.

La mayoría de los investigadores creen que los tumores se desarrollan a partir de una sola célula transformada. Por tanto, subpoblaciones que ostenten propiedades diferentes podrán surgir sólo en el caso de que varias células de un tumor evolucionen de un modo independiente, aunque se hallen en el mismo entorno. Un tumor de lento crecimiento en pacientes humanos (o la propagación repetida de tumores por trasplantes consecutivos en animales de experimentación) suele conducir a cambios progresivos en la estructura, propiedades antigénicas, características metabólicas, propiedades del crecimiento y potencia metastásica del tumor. Varios expertos han llamado la atención sobre este fenómeno de la progresividad del tumor; entre ellos debemos enumerar a Peter C. Nowell, de la Facultad de Medicina de la Universidad de Pennsylvania, Leslie Foulds, del Chester Beatty Research Institute londinense, George y Eva Klein, del Instituto Carolino de Estocolmo, Emmanuel Farber, de la Universidad de Toronto, y H. J.– P. Ryser, de la Facultad de Medicina de la Universidad de Boston. Conforme una población de cé-



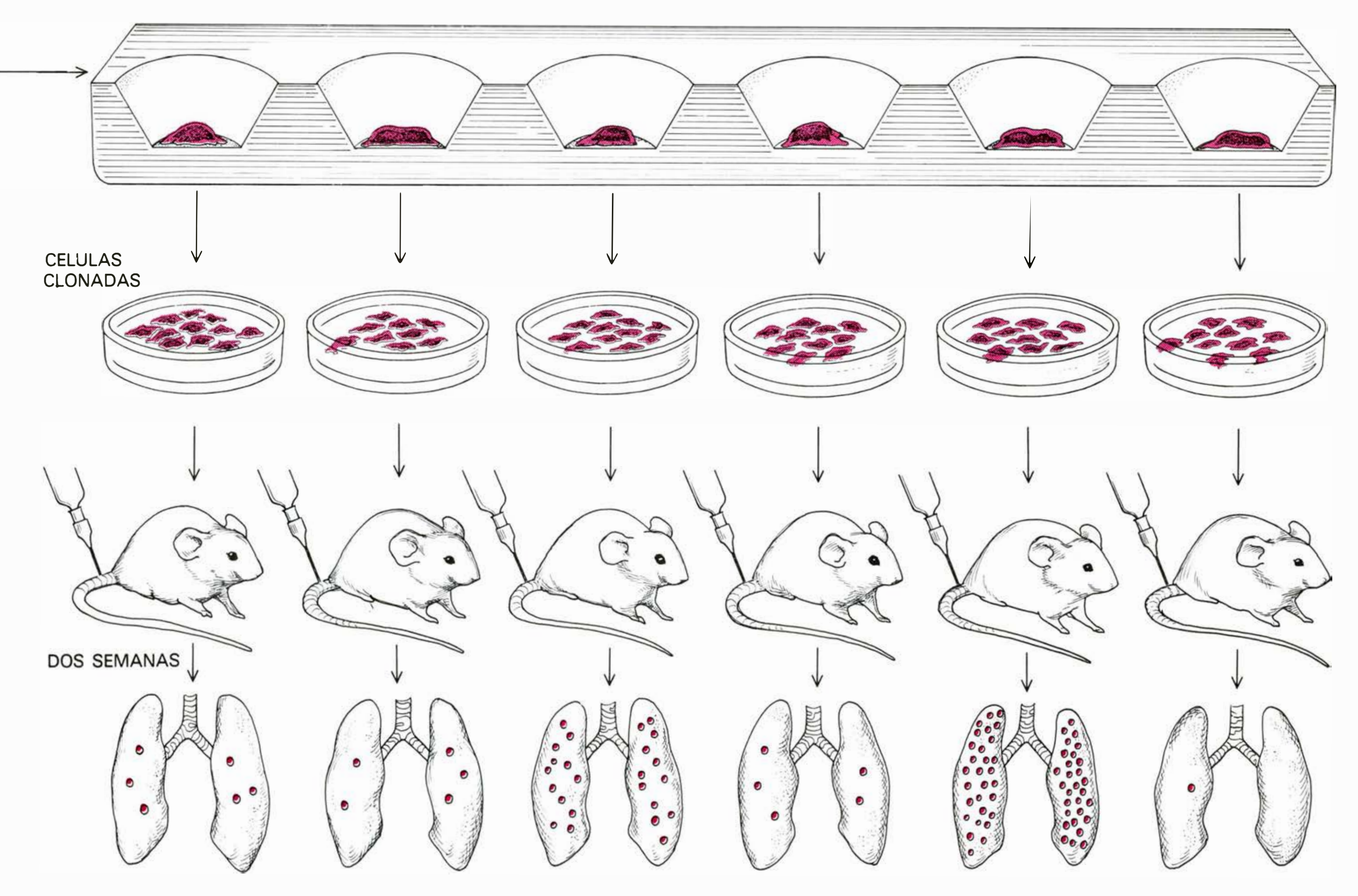
HETEROGENEIDAD METASTASICA de una población de células tumorales, demostrada a través de los experimentos realizados en el laboratorio del autor y en el de Fidler. En cada caso, una población celular tumoral no seleccionada (*1*) se dividió en dos partes. Una (*2*) se inyectó por vía intravenosa en muchos grupos de ratones. La otra (*3*) se empleó para producir un gran número de clones, descendiente cada uno de una sola célula;

lulas tumorales va aumentando poco a poco, o bien cambia todo el tumor o bien parecen surgir subpoblaciones portadoras de propiedades distintas, aumentando frecuentemente en la población por presión selectiva de las defensas del hospedante, resultando en última instancia la formación de células de malignidad más acusada. ¿Existen subpoblaciones que ostentan diferentes potencias metastásicas ya en el tumor primitivo? ¿O acaso se trata de una simple adaptación a un entorno dado por parte de toda la población?

Un método de distinguir entre ambas posibilidades es subsidiario del experimento, ya clásico, de Salvador Luria y Max Delbrück. En 1943 ambos autores pudieron distinguir por primera vez entre selección y adaptación como origen de líneas mutantes de bacterias. Fidler y Margaret L. Kripke repitieron el experimento con la línea melanómica *B*16 no seleccionada y con una línea del fibrosarcoma (tumor del tejido conjuntivo) procedente de un tumor inducido en un ratón por irradiación ultravioleta; nosotros realizamos el experimento en Irvine con una línea de sarcoma de ratón. En cada caso, las suspensiones de células tumorales no seleccionadas se dividían en dos partes. Una se inyectó por vía intravenosa en un grupo nutrido de ratones. La otra parte se dividió en cultivos independientes, de cada uno de los cuales se aisló una sola célula que creció hasta producir un gran número de clones distintos, o subpoblaciones de células con una constitución genética idéntica. Estos clones se inyectaron también por vía intravenosa en grupos de ratones.

Si el número de colonias tumorales en los pulmones de ratón que recibieron varias subpoblaciones de clones resultaban parecidas en número al de colonias de ratones que recibieron la línea no clonada, podría deducirse que la línea tumoral originaria era homogénea. Si diferentes subpoblaciones de clones daban origen a un gran número de colonias distintas, en comparación con la línea no clonada, podría entenderse que la línea no seleccionada original era heterogénea, con células de baja y alta potencia metastásica preexistente en ellas. En los tres experimentos se llegó al último resultado. Ello implica que las poblaciones celulares tumorales son altamente heterogéneas en lo que se refiere a las propiedades que regulan la malignidad, y, también, que en un tumor hay escasas subpoblaciones altamente metastásicas antes de que ocurra el proceso de progresión y selección tumoral de poblaciones celulares con propiedades metastásicas acusadas.

Cuando una célula normal se transforma en una célula tumoral, la superficie celular muestra nuevos (o al menos modificados) antígenos, que son estructuras moleculares que el sistema inmunitario identifica como extrañas (véase "Inmunología del cáncer", por Lloyd J. Old; INVESTIGACIÓN Y CIENCIA, julio, 1977). La aparición de estos antígenos, asociados al tumor, puede estimular en el hospedante una respuesta inmune que conduce a la intercepción y destrucción de las células tumorales. F. M. Burnet, del Walter and Eliza Hall Institute de investigación clínica de Melbourne, propuso que las células normales se transforman periódicamente, pero un sistema generalizado de "vigilancia inmunológica" evita que formen tumo-

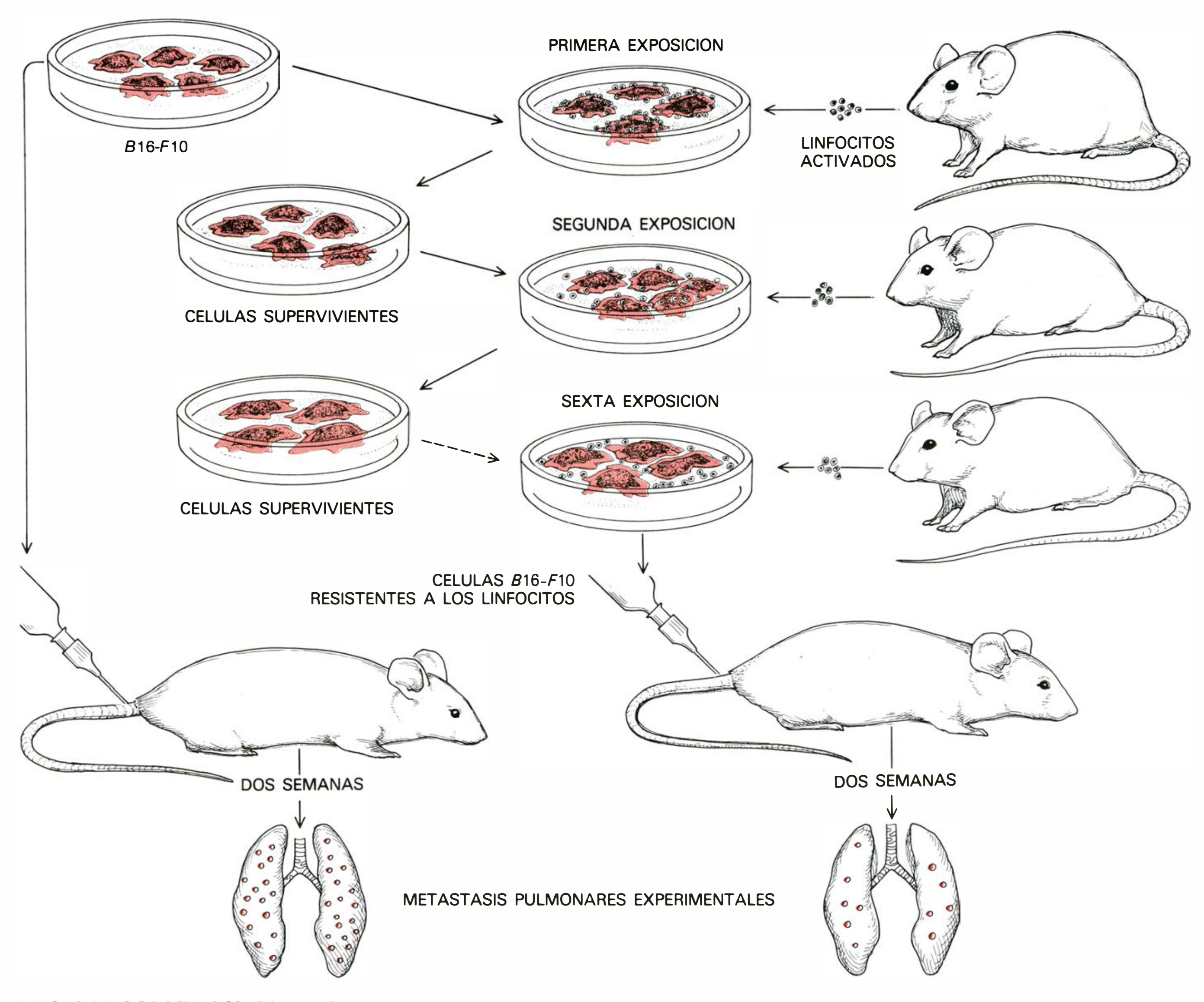


cada clon unicelular creció, y las células clonadas se inyectaron en otros grupos de ratones. La cantidad de metástasis producidas por las células no clonadas y por cada una de las pertenecientes a los clones resultó ser muy distinta. Ello evidenciaba que la población original (*1*) tenía que haber sido muy heterogénea. De haber sido homogénea, el número de metástasis producidas por células no clonadas y por cada clon hubiera sido similar.

res; por tanto, los escasos tumores que se presentan se deben a células ocasionalmente transformadas que se las ingenian para eludir el sistema de control. La eficacia del sistema inmunológico en lo referente a la protección contra la manifestación y posterior crecimiento progresivo de los tumores ha sido cuestionada por otros investigadores. Richmond T. Prehn, del Jackson Laboratory, y Osias Stutman, del Memorial Sloan-Kettering Cancer Center, están entre los que dudan que el sistema generalizado de vigilancia inmunológica constituya un medio eficaz de defensa antitumoral.

Ciertamente, la respuesta inmunológica incide en el progreso y metástasis de determinados cánceres humanos y en tumores animales. Y así, investigadores como Robert W. Baldwin, de la Universidad de Nottingham, Luka Milas, del Instituto yugoslavo Rudjer Bošković, Vilas V. Likhite, de la Facultad de Medicina de Harvard, y John E. Castro, de la Royal Postgraduate Medical School londinense, han demostrado que la metástasis de ciertos tumores puede inhibirse si se estimula el sistema inmunitario a través de la introducción de ciertos microorganismos, como el bacilo de Calmette-Guérin (BCG) y *Corynebacterium parvum.* No se comprende bien el mecanismo de dicha estimulación; quizás obedezca a una estimulación no específica del sistema inmunitario, o bien a una respuesta específica a los antígenos de la superficie del microorganismo, que son parecidos a los antígenos de las células tumorales.

Pero el control de la metástasis no se corresponde siempre con la fuerza del sistema inmunológico. Los experimentos fracasan a menudo a la hora de demostrar que la metástasis producida por vía sanguínea sea más relevante en los animales cuyo sistema inmunológico se ha debilitado de un modo experimental que en los animales cuyo sistema inmunológico permanecía inalterado. Es más, el aumento de defensas inmunitarias en ciertos pacientes y en algunos animales de experimentación provoca un potenciamiento de la metástasis. Cuando los Fisher rebajaron la inmunidad antitumoral del ratón, mediante la inyección de anticuerpos que destruyeran la respuesta de los linfocitos (que constituyen un importante componente celular del sistema inmunológico) a las células tumorales, observaron que no se formaban más, sino menos, metástasis. Por otro lado, Michael Feldman, del Weizmann Institute de Israel, advirtió que la inyección, por vía intravenosa, de grandes cantidades de linfocitos previamente activados para acabar con las células del



EFECTO DE LOS LINFOCITOS sobre la metástasis, estudiado por Fidler. Este expuso células del melanoma *B*16-*F*10 a linfocitos activados, es decir, linfocitos procedentes de los ganglios linfáticos y del bazo de ratones que se mostraban inmunes a células tumorales *B*16. Algunas de las células *B*16-*F*10 murieron; las supervivientes fueron sometidas de nuevo a los linfocitos activados. Tras repetir por seis veces el ciclo, las células *B*16-*F*10 remanentes se tornaron completamente resistentes a ser destruidas por dichos linfocitos. Pero cuando se inyectaron las células resistentes en ratones, producían menos metástasis pulmonares que las células *B*16-*F*10 sin tratar. La resistencia a los linfocitos no contribuyó a que las células metastatizaran. SE se redujo su potencial metastásico al eliminar las interacciones célula tumoral-linfocitos, importantes para la detención y la supervivencia.

carcinoma, podían inhibir la metástasis de dicho carcinoma en animales de experimentación.

Estos resultados, que se mostraban contradictorios, indujeron a Prehn a sugerir que dos sistemas inmunológicos competitivos podían verse involucrados: uno con la responsabilidad de matar los tumores y el otro para estimular a los mismos. Prehn descubrió que si introducía una pequeña cantidad de linfocitos antitumorales en ratones resistentes al tumor, estimulaba el crecimiento tumoral, en tanto que la inyección de una cantidad considerable de linfocitos inhibía el crecimiento tumoral. Se han realizado experimentos similares con el sistema melanómico *B*16. La inyección en ratones de un pequeño número de linfocitos mezclados con células *B*16 trajo consigo un incremento significativo en las metástasis pulmonares, pero cuando lo que se mezclaba era grandes proporciones de linfocitos con células tumorales se registraban muy pocas metástasis.

¿Cómo podría un número reducido de linfocitos estimular la detención y supervivencia de células melanómicas en el torrente sanguíneo? La respuesta dista mucho de ser evidente, si bien una observación oportuna pudiera ser que, cuando se mezclan en una suspensión linfocitos y células del melanoma *B*16, se congregan en pequeños grumos. Como expuse antes, la formación de concreciones aumenta la detención y supervivencia de células tumorales. Ello pudiera deberse sencillamente a que las concreciones se alojan en los capilares pulmonares perfectamente, o bien porque un grupo de substancias llamadas linfoquinas, que pueden liberar los linfocitos tras su interactuación con células tumorales, ejercen cierto efecto favorable sobre la supervivencia de células malignas.

Entonces, ¿por qué la inyección de grandes cantidades de linfocitos inhiben la metástasis? Una proporción notable de linfocitos se muestra más eficaz que un número reducido de ellos a la hora de formar agregados con células *B*16; lo cual podría, presumiblemente, facilitar la detención e implantación de las células tumorales. Por otro lado, grandes cantidades de linfocitos inhiben la metástasis al matar células malignas. Pudiera ocurrir que el efecto antimetastásico de grandes cantidades de linfocitos se deba realmente a la presencia, entre ellos, de una corta subpoblación de linfocitos "asesinos" que han sido activados específicamente para atacar a las células tumorales. O fuera quizá que los linfocitos quedaran contaminados a lo largo de los distintos experimentos por unas cuantas células con diferente función inmunitaria que eran las encargadas de destruir las células tumorales.

En un esfuerzo por averiguar si los linfocitos eran o no los responsables directos de la inhibición de la metástasis, Fidler cultivó células del melanoma *B*16-*F*10 con linfocitos activados. La mayoría de las células tumorales murieron, pero algunas pocas sobrevivieron. Estas últimas volvieron a ser sometidas a cultivo y expuestas a la acción de los linfocitos de nuevo; una vez más, las supervivientes fueron aisladas, cultivadas y mezcladas con linfocitos activados. Tras seis selecciones como la precedente para la resistencia al ataque de los linfocitos, las células *B*16-*F*10 vivas resultaron completamente resistentes al ataque destructor de los linfocitos. Dichas células resistentes volvieron a inyectarse por vía intravenosa en ratones. Formaron menos metástasis que las células *F*10 no tratadas; en realidad, no más metástasis que las *B*16-*F*1 altamente metastásicas. Además, ni siquiera formaron concreciones cuando se las mezcló en una suspensión con linfocitos. Fidler concluyó que el efecto usual de los linfocitos en las células *B*16 estriba en agruparlas en la circulación y facilitar, por consiguiente, la metástasis, y que la inhibición de la metástasis observada cuando existe una elevada proporción de linfocitos/células tumorales viene causada probablemente por otro tipo de células inmunológicas que se halla junto con los linfocitos.

Esta conclusión se apoya en experimentos con ratones "desnudos", línea de ratones sin pelo que carecen de un timo maduro y, por consiguiente, muestran una inmunidad disminuida; así, por ejemplo, no rechazan injertos de piel de otras especies. Cuando inyectamos células *B*16 marcadas en ratones desnudos, las células se detenían y sobrevivían en una proporción más baja que ocurría en el caso de ratones con un sistema inmunitario normal. Un inyección de linfocitos procedentes de ratones normales en ratones desnudos un día antes de la inoculación de la célula tumoral restauraba parcialmente la configuración normal de metástasis. La moraleja, de nuevo, es que, por lo menos en el sistema *B*16, los linfocitos estimulan metástasis experimentales.

Es verdad que en otros experimentos, entre los que se cuentan los realizados en los laboratorios de Feldman y Baldwin, así como en el de Peter Alexander, integrado en el Chester Beatty Research, las respuestas inmunológicas mediatizadas por linfocitos han mostrado su eficacia en la prevención de la diseminación metastásica o en la eliminación de micrometástasis establecidas ya. El hecho de que no se pueda predecir fácilmente si la inmunidad del hospedante potenciará o inhibirá la diseminación tumoral significa que la inmunidad y la metástasis no están relacionadas de una manera inmediata. Los hallazgos experimentales se ven corroborados por observaciones clínicas: los procedimientos que se diseñan para aumentar el estatuto inmunológico de un paciente inhiben la metástasis en algunos casos y en otros la estimulan. La quimioterapia o la inmunoterapia del cáncer pueden tender a seleccionar subpoblaciones de células resistentes altamente metastásicas, además de destruir algunas células cancerosas. Habrá que discutir en cada caso si el efecto terapéutico predomina sobre el efecto de selección indeseable.

A la hora de esbozar una terapéutica eficaz habría que tener en cuenta, ciertamente, que sólo unas, pocas, células altamente metastásicas de una población tumoral pueden acabar por producir los mayores daños. Experimentos recientes que Reuben Lotan ha llevado a cabo en mi laboratorio muestran que cada línea seleccionada del melanoma *B*16 o del clon *B*16 no seleccionado responde de un modo distinto a las drogas que le administran por quimioterapia. En la heterogeneidad de las poblaciones tumorales puede residir la causa de ciertos fallos clínicos y experimentales de las drogas quimioterapéuticas si las subpoblaciones resistentes logran sobrevivir al tratamiento de fármacos. Puede elegirse una droga porque inhibe el desarrollo de un tumor primario o de su correspondiente en un cultivo tisular y, a pesar de todo, puede fallar en la detención del crecimiento de algunas subpoblaciones de células tumorales que se hallan presentes en las micrometástasis.

Todos los ensayos que habrán de acometerse en el futuro se centrarán, sin lugar a dudas, en la comprensión de la biología del avance del tumor y el proceso metastásico. Las investigaciones sobre líneas de células tumorales que metastizan preferentemente en órganos determinados ayudarán a desvelar los mecanismos que controlan el asentamiento de las mismas y su implantación y desarrollo en cierto ambiente; los estudios sobre líneas celulares que son resistentes a ciertas drogas o al sistema inmunológico habrán de permitir conocer por qué algunos tumores escapan de la terapia. Saber cada día más sobre la metástasis y progresión tumoral en los sistemas tumorales animales nos llevará al establecimiento de una terapéutica más racional para el tratamiento eficaz del cáncer.

El espectro del hidrógeno atómico

Durante casi un siglo, la luz emitida por los átomos más sencillos ha sido la principal base experimental de las teorías sobre la estructura de la materia. Con la ayuda de los láseres se continúa explorando el espectro del hidrógeno

Theodor W. Hänsch, Arthur L. Schawlow y George W. Series

El espectro del átomo de hidrógeno ha resultado ser la piedra de Rosetta de la física moderna: cuando se logró descifrar esta distribución de líneas, se entendieron muchas otras cosas. En particular, fue en gran parte el esfuerzo para explicar el espectro de la luz emitida por el átomo de hidrógeno lo que inspiró las leyes de la mecánica cuántica. Desde entonces se ha observado que estas leyes se aplican no sólo al átomo de hidrógeno sino también a otros átomos, a las moléculas y a la materia en general. Constituyen los fundamentos de la química moderna, de la física del estado sólido y aún de ciencias aplicadas tales como la electrónica.

La posición central que ocupa el átomo de hidrógeno en la historia de la física del siglo XX pudiera inducir a pensar que el espectro se conoce en todos sus detalles desde hace tiempo. Esto no es cierto. Sólo en los últimos años se han resuelto algunas de las características más sutiles del espectro, y muchas otras no se han observado todavía directamente. La medición de las posiciones de las rayas espectrales continúa siendo hoy en día una prueba importante de las predicciones de la teoría cuántica.

El espectro de rayas

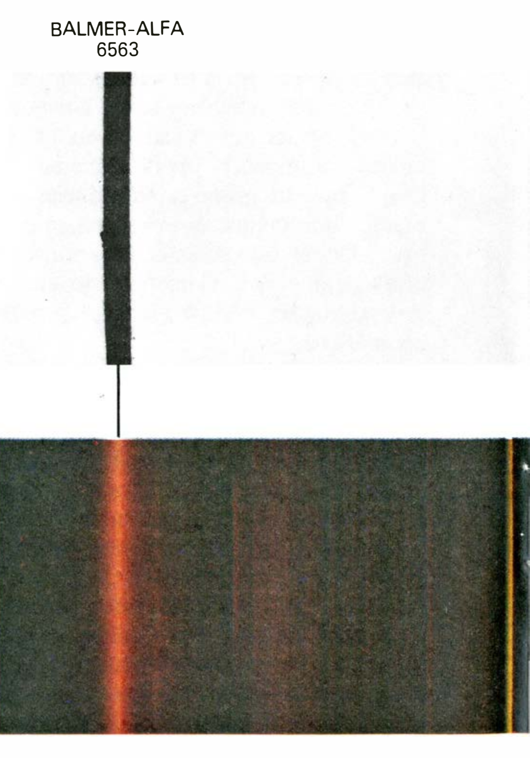
Los descubrimientos más recientes en el análisis del espectro pueden atribuirse a una nueva herramienta de la espectroscopía: el láser. Como los detalles más finos del espectro están densamente apretados sólo pueden distinguirse usando luz que sea altamente monocromática, o confinada a una banda muy estrecha de longitudes de ondas. El láser es una fuente luminosa de este tipo. Como resultado de técnicas de laboratorio ingeniosas, el láser se ha convertido en un instrumento espectroscópico de resolución sin precedentes.

Cuando la luz emitida por un filamento caliente es dispersada, de acuerdo con su longitud de onda, por un prisma o una red de difracción, resulta un abanico continuo de colores. El espectro de un gas puro y rarificado de átomos o moléculas consiste, sin embargo, en un conjunto de rayas discretas. Si se registra el espectro de luz emitido por el gas, aparece como una secuencia de rayas brillantes sobre un fondo obscuro. Si la luz pasa a través de un gas, quedan absorbidas longitudes de onda discreta, dando origen a rayas obscuras sobre un fondo brillante.

El hidrógeno es el más sencillo de los átomos, al estar constituido por un solo electrón y un núcleo formado por un solo protón; por tanto, cabe esperar que tenga el espectro más simple. Pero no se crea que dicho espectro resulta fácil de obtener. La raya más prominente fue detectada en 1853 por Anders Jonas Angstrom. (La unidad usual en la que se miden las longitudes de onda de la luz se llama angstrom, en su honor; y un angstrom equivale a 10^{-8} centímetros.) En las dos décadas siguientes se observaron tres nuevas rayas, si bien las primera serie extensa de rayas del hidrógeno atómico no se encontró en el laboratorio sino en los espectros de las estrellas. En 1881, estudiando las primeras fotografías de espectros estelares, Sir William Huggins identificó 10 rayas como producidas por emisión de hidrógeno atómico.

Quizá parezca sorprendente que rayas del espectro del hidrógeno se vieran antes en observaciones astronómicas que en los experimentos terrestres. La dificultad que entraña la medición del espectro en el laboratorio no estriba en la detección de las rayas sino en la preparación de hidrógeno atómico puro. El hidrógeno gaseoso ordinario consta de moléculas diatómicas (H_2), que tienen un espectro mucho más complicado que el espectro del átomo de hidrógeno aislado. Dividir las moléculas requiere más energía de la que pueden suministrar la mayoría de las excitaciones térmicas, tales como una llama. Un aparato que aporta energía de forma muy eficaz es el tubo de descarga gaseosa, en el que los electrones acelerados por la diferencia de potencial aplicada disocian las moléculas. A lo largo de los años se han descubierto medios para resaltar el espectro atómico y suprimir el molecular.

Las rayas espectrales detectadas por Huggins tienen longitudes de onda distribuidas desde la parte roja del espectro visible hasta el ultravioleta próximo. Las



CONJUNTO DE RAYAS DISTINTAS que se forman cuando la luz emitida por átomos de hidrógeno es dispersada según su longitud de onda. Las rayas son de la serie de Balmer y sus longitudes de onda aproximadas se dan en angstrom (el angstrom es una unidad de longitud que equivale a 10^{-8} centímetros). Esta fotografía en color fue hecha por Jon Brenneis y uno de los autores

primeras dos rayas están netamente separadas, pero las subsiguientes se hallan separadas a través de intervalos que decrecen suavemente; las que poseen longitudes de onda más cortas están muy juntas entre sí. En 1885, trabajando sobre las medidas astronómicas, Johann Jakob Balmer descubrió que se podían explicar las posiciones de todas las rayas conocidas aplicando una fórmula empírica sencilla. Todo este conjunto de rayas se conoce con el nombre de serie de Balmer. Otro grupo de rayas, la serie de Lyman, yace en el ultravioleta lejano, y existen otras series de rayas con longitudes de onda más largas. En el seno de cada serie, las distintas rayas se designan con letras griegas, a partir de la línea de longitud de onda más larga. En efecto, la raya brillante roja vista por primera vez por Angstrom es la raya Balmer-alfa; la siguiente, la raya Balmer-beta, y así sucesivamente. A causa de la prominencia de la raya Balmer-alfa se la llama a veces simplemente la raya hidrógeno-alfa o H-alfa.

En 1889, Johannes Rydberg descubrió que las rayas del espectro de muchos metales, en particular las de los metales alcalinos, podían ajustarse mediante una fórmula empírica sencilla. Una versión posterior de la fórmula, que es equivalente a la ecuación de Rydberg pero más explícita, tiene la forma

$$\frac{1}{\lambda} = R\left[\frac{1}{(m+b)^2} - \frac{1}{(n+c)^2}\right].$$

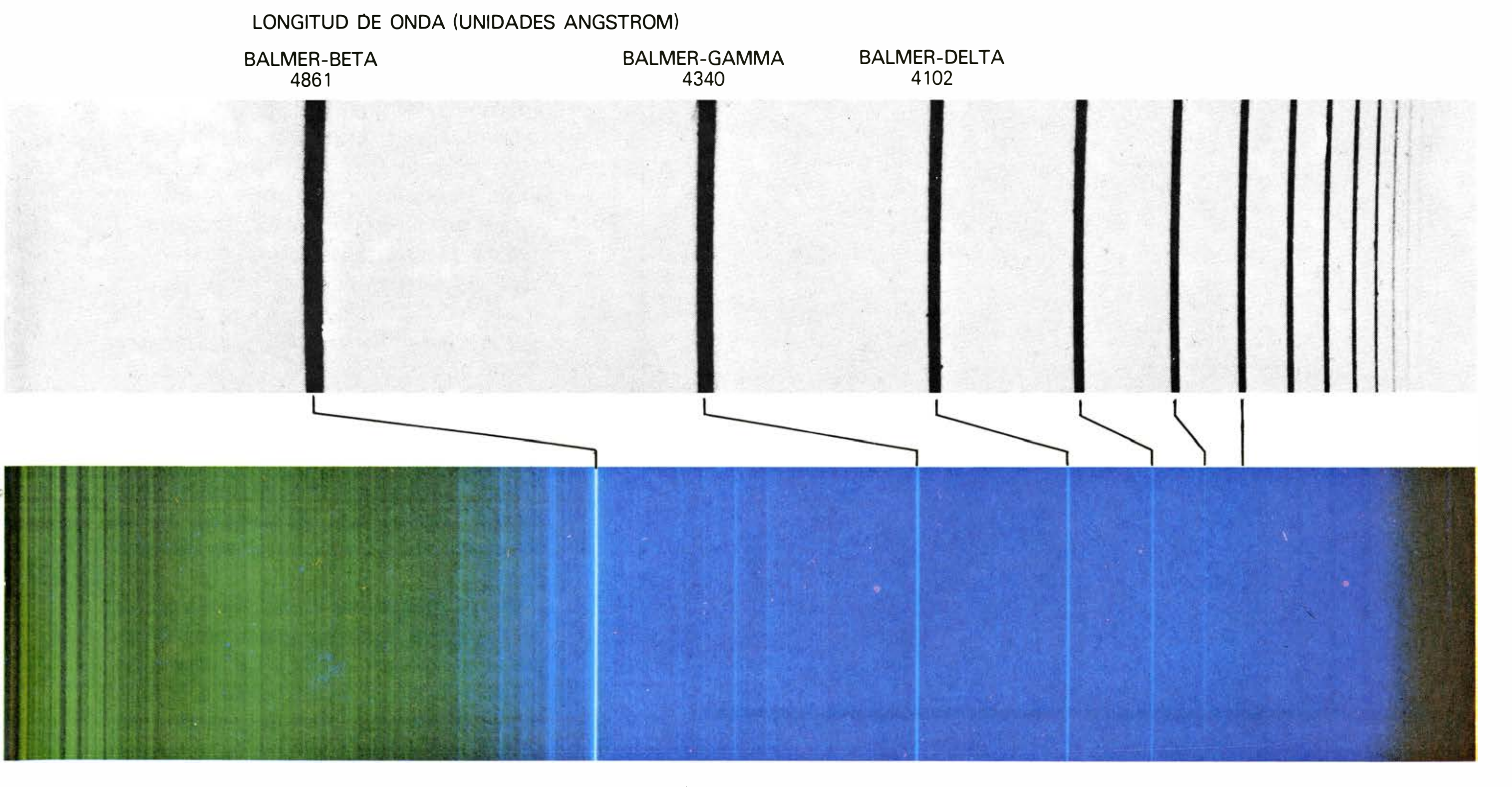
Aquí λ (la letra griega lambda) simboliza la longitud de onda de una determinada raya en un espectro atómico; m y n son números que toman valores enteros sucesivos (1, 2, 3, etcétera) y R, b y c son constantes. Los valores de b y c dependen de qué serie de rayas se esté midiendo, pero R es la misma para todas las rayas de todos los elementos. La R se denomina hoy constante de Rydberg; si la longitud de onda se expresa en metros, R tiene el valor $1{,}097 \times 10^7$ y las dimensiones de recíproco de metro. Después de que Rydberg comprobara su fórmula con un cierto número de espectros escribió: "Acababa de comprobar varias formas de la función cuando conocí la comunicación del señor Balmer sobre las rayas espectrales del átomo de hidrógeno. Quedé muy satisfecho al hallar que su fórmula es un caso especial de la mía, con el mismo valor de R y con $c = 0$". Por tanto, para el átomo de hidrógeno la ecuación tiene una forma más simple:

$$\frac{1}{\lambda} = R\left[\frac{1}{m^2} - \frac{1}{n^2}\right].$$

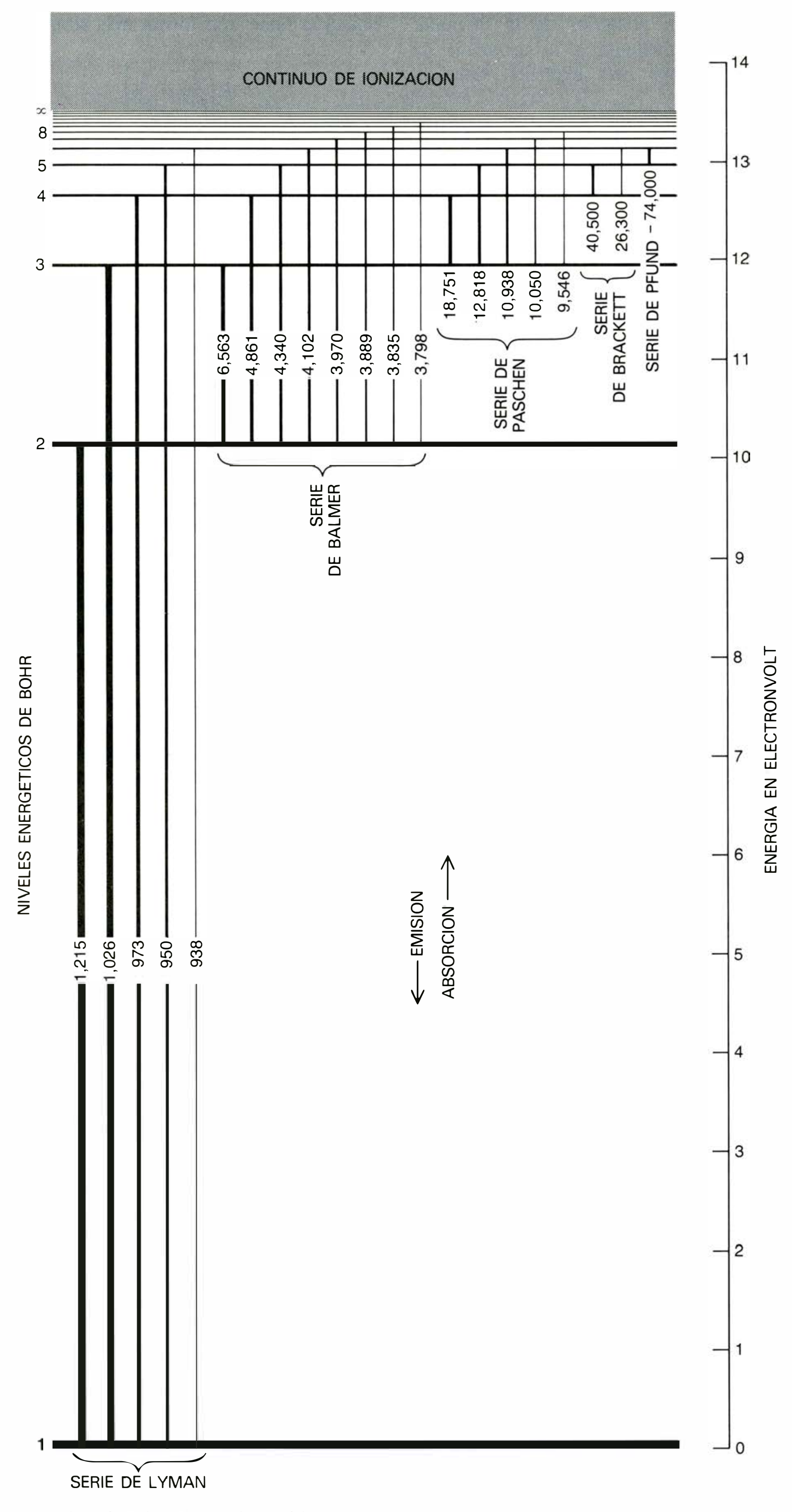
Cuando se sustituyen en esta fórmula valores apropiados de m y n, se obtienen las longitudes de onda de todas las rayas del espectro del hidrógeno.

Estructura atómica

Niels Bohr pasó la mayor parte del año 1912 en la Universidad de Manchester trabajando en el laboratorio de Ernest Rutherford, quien acababa de realizar una contribución vital para la comprensión de la estructura atómica. Por aquel entonces, las fórmulas de Balmer y Rydberg eran bien conocidas por los espectroscopistas, y habían sido ampliamente comprobadas; pero Bohr no había oído hablar nunca de ellas. Cuando se enteró de la fórmula de Balmer, la pudo deducir casi inmediatamente de las propiedades del átomo de hidrógeno. Abraham Pais y T. D. Lee cuentan cómo andando el tiempo se preguntó a Bohr a qué se debía que no oyera hablar de la fórmula de Balmer. En respuesta contestó que en aquellos



(Schawlow) en el laboratorio de este último en la Universidad de Stanford. El espectro en blanco y negro fue obtenido en 1927 por Gerhard Herzberg, que estaba entonces en la Universidad Técnica de Darmstadt, Alemania. Los intervalos entre las rayas son distintos en las dos fotografías, pues Brenneis y Schawlow dispersaron la luz mediante una red de difracción, mientras que Herzberg empleó un prisma; el prisma introduce una dispersión porque el índice de refracción de cristal no es función lineal de la longitud de onda. En ambos casos, la fuente del espectro era un tubo de descarga gaseosa en el que las moléculas de hidrógeno (H_2) eran disociadas en átomos mediante un potencial eléctrico. En la fotografía en color, las bandas amarillas y verdes y el fondo continuo azul de la derecha no son parte del espectro del hidrógeno atómico sino que se deben a los contaminantes moleculares en el tubo de descarga. Herzberg eliminó esta contaminación enfriando el tubo en nitrógeno líquido. Las rayas a la derecha de la Balmer-delta, que aparecen blancas o ligeramente azuladas en la fotografía en color, están en realidad en la región ultravioleta del espectro electromagnético.



TRANSICIONES ENTRE ESTADOS del átomo de hidrógeno. Tales transiciones dan origen al espectro de rayas. Cada estado se indica mediante un número entero, llamado el número cuántico principal del correspondiente nivel energético de Bohr. Cuando el átomo cambia de un estado a otro, la diferencia de energía aparece como un cuanto de radiación. La energía del cuanto es directamente proporcional a la frecuencia de la radiación e inversamente proporcional a su longitud de onda. La absorción de radiación estimula una transición a un estado de energía más alta; un átomo que cae a un estado de menor energía emite radiación. El espectro se organiza en una serie de rayas que comparten el nivel más bajo. Las longitudes de onda se dan en angstrom; la intensidad relativa se indica por su espesor.

años muchos físicos consideraban los espectros atómicos sumamente complicados como para que formaran parte de la física fundamental. Se les consideraba, puntualizó, como las notas de un piano, cuyos tonos dependían de una forma complicada de la estructura del instrumento como un todo.

La referencia a las notas musicales no era casual. Durante el siglo XIX muchos investigadores habían especulado que la luz emitida por los átomos podía entenderse a través de cierta analogía con los modos de vibración de un cuerpo sólido. La presencia de múltiples rayas en cada serie podría explicarse entonces suponiendo que la primera raya constituía el modo fundamental y el resto eran sobretonos, o armónicos. Pero ningún espectro atómico pudo resolverse en una frecuencia fundamental y sus sobretonos, y, por tanto, hubo que acometer un examen más directo del mismo átomo.

La contribución de Rutherford fue la demostración de que el átomo no era un cuerpo sólido, sino que consta de un núcleo pequeño y muy denso cargado positivamente y de un número adecuado de electrones, que están cargados negativamente, para formar un sistema estable y eléctricamente neutro. Bohr intentó explicar el espectro de los átomos en el contexto de este modelo. Al hacerlo, se expuso al ridículo de sus contemporáneos, pues avanzó la hipótesis de que las leyes establecidas de la física no eran aplicables dentro del átomo. Dichas leyes predecían que cualquier electrón ligado radiaría toda su energía y caería sobre el núcleo. Bohr propuso que un átomo podría existir en un estado en el que la energía del electrón no se disipara: un estado estacionario. Se emite o absorbe luz, sugirió, sólo cuando un electrón pasa de un estado estacionario a otro. La frecuencia de la luz viene determinada por la diferencia de energía entre los dos estados, de acuerdo con el principio cuántico introducido por Max Planck. La frecuencia se encuentra dividiendo la diferencia de energías por la constante de Planck, que se designa por h.

En el modelo de Bohr, las fórmulas de Balmer y Rydberg tienen una interpretación física obvia. La expresión $(1/m^2 - 1/n^2)$ es proporcional a la diferencia de energía entre dos estados del átomo: los enteros m y n caracterizan cada uno de los estados. Bohr hizo algo más que explicar esta ecuación empírica; evaluó la constante de Rydberg en términos de tres cantidades fundamentales: la carga eléctrica del electrón e, la masa del electrón m y la constante de Planck h. Si, por sencillez, el núcleo de un

átomo se supone con masa infinita, entonces la constante de Rydberg viene dada por la ecuación

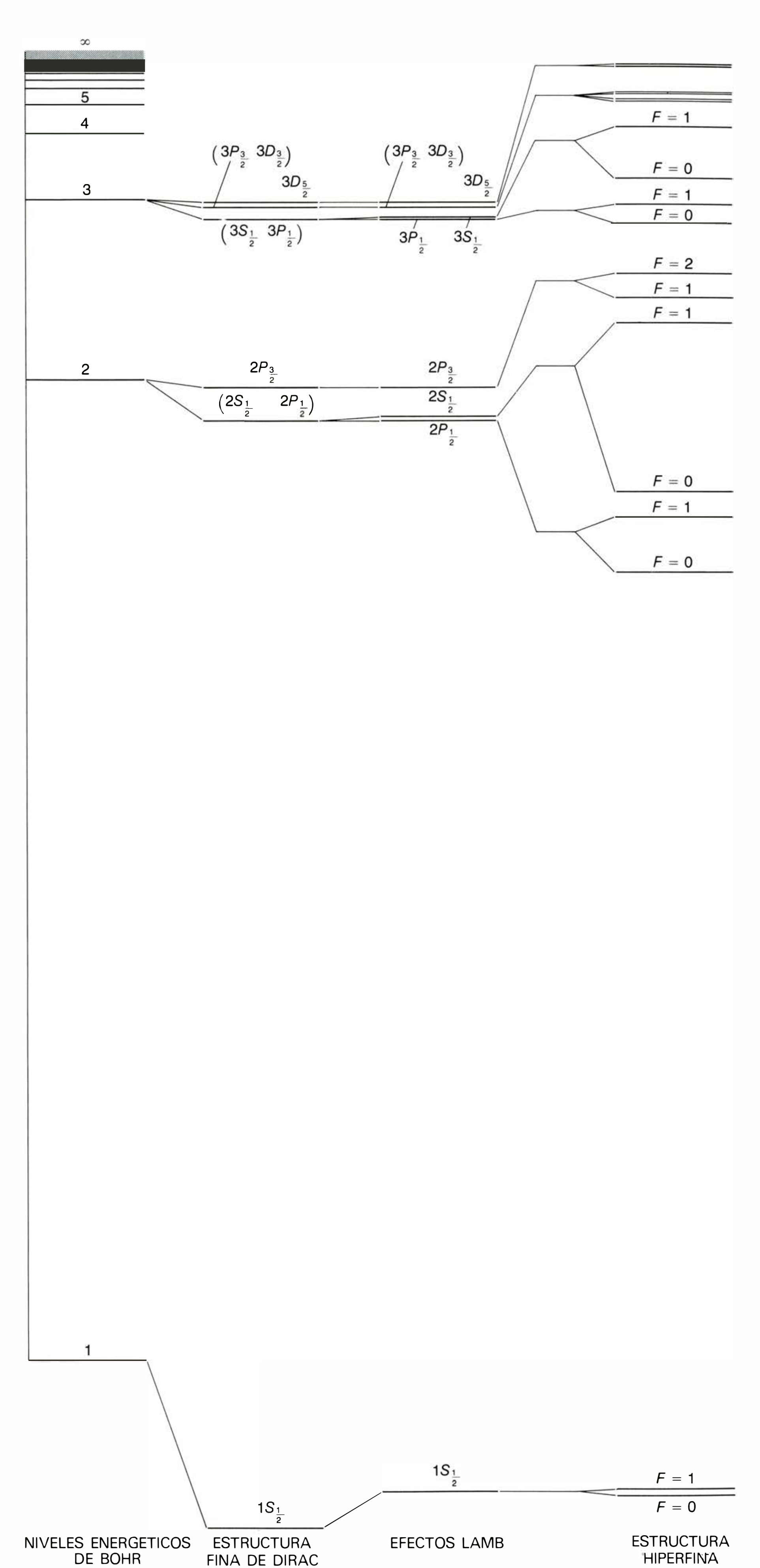
$$R = \frac{2\pi^2 me^4}{h^2}$$

Posteriores mejoras han complicado la fórmula empírica de Rydberg para las longitudes de onda de las rayas espectrales; por cuyo motivo la constante de Rydberg se define hoy como la combinación de m, e y h.

Bohr imaginó que los electrones de un átomo siguen órbitas circulares. Al calcular la energía de los estados estacionarios, empezó suponiendo que, para las órbitas de un diámetro muy grande, la frecuencia de la luz emitida debería corresponder a la frecuencia del electrón en esta órbita, que era lo que predecía la física clásica. Este "principio de correspondencia" conducía a la conclusión intrigante de que, para un estado estacionario designado por el entero n, el momento angular del electrón era igual a $nh/2\pi$. Por tanto, el momento angular, al igual que la energía, podían cambiar sólo de forma discreta. Esta "cuantización" del momento angular es de importancia fundamental: sin ella, el átomo podría radiar en cualquier frecuencia y el estado ya no sería estacionario.

Mecánica cuántica

El átomo de Bohr constituyó un gran avance conceptual sobre todas las teorías anteriores, pero pronto se vio que resultaba inadecuado para explicar las características observadas en los espectros atómicos, aún aquellas del espectro más simple, el del átomo de hidrógeno. En efecto, un rasgo del espectro del hidrógeno que la teoría de Bohr no podía explicar era conocido desde hacia 20 años cuando se construyó el modelo. En 1892, A. A. Michelson había empleado el interferómetro que él había inventado



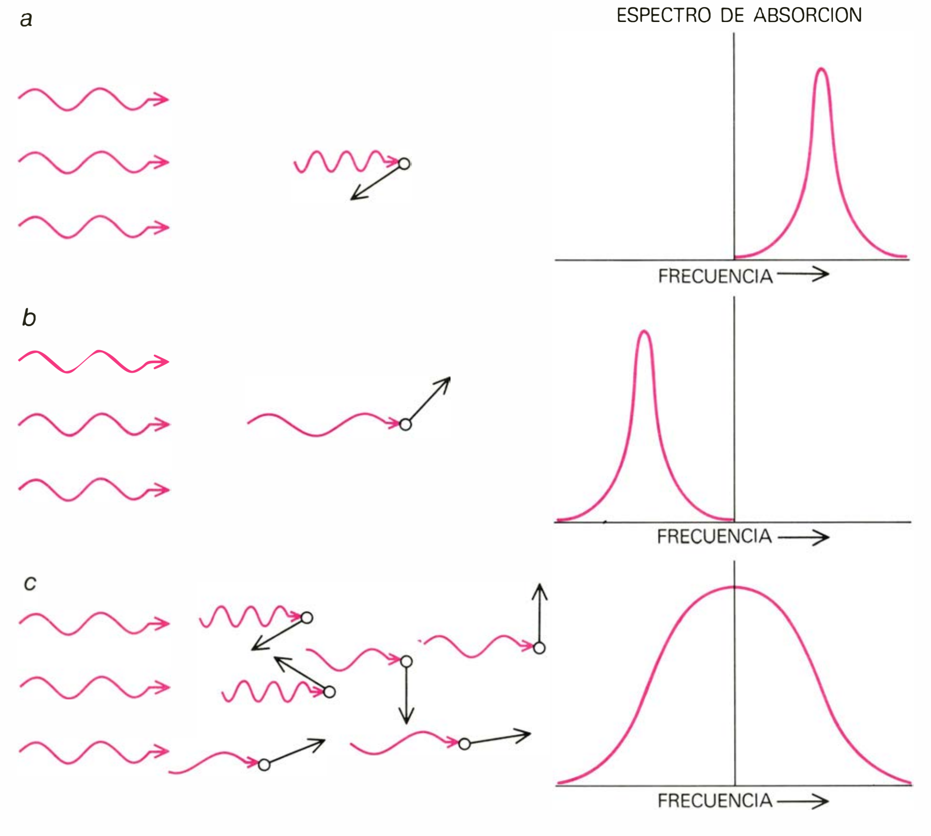
DESDOBLAMIENTO de los niveles energéticos, debido principalmente a interacciones relativistas y magnéticas relacionadas con el momento angular del átomo. El único electrón de un átomo de hidrógeno puede tener tanto momento angular orbital como de spin, y las distintas combinaciones posibles de estas dos cantidades tienen generalmente energías distintas. En razón de lo cual, los niveles de energía Bohr se dividen en componentes, que se denominan niveles de estructura fina. Un sutil efecto, llamado efecto Lamb, desplaza ciertas componentes de la estructura fina, creando nuevos desdoblamientos. Por último, la interacción del momento magnético del electrón con el del núcleo da origen a la estructura hiperfina. La estructura fina y el efecto Lamb se representan aquí a una escala mayor que los niveles de Bohr y los desdoblamientos hiperfinos de segundo y tercer niveles de Bohr se representan a una escala aún mayor.

para examinar las formas de las distintas rayas del espectro, esto es, la variación de su intensidad como función de la longitud de onda. Halló que la raya Balmer-alfa no era una sola raya, sino que tenía dos componentes separadas entre sí por una longitud de onda de 0,14 angstrom.

Bohr sugirió que el desdoblamiento de la raya Balmer-alfa podía explicarse si la órbita del electrón en el átomo de hidrógeno no era circular sino elíptica. Una corrección al movimiento orbital exigida por la teoría de la relatividad restringida dividiría cada estado estacionario en un grupo de estados. Un análisis más completo de la idea fue presentado posteriormente por Arnold Sommerfeld, quien dedujo el desdoblamiento observado a partir de las propiedades calculadas de las órbitas. En su cálculo, introdujo un importante número sin dimensiones que se llama la constante de estructura fina y es igual a $2\pi e^2/hc$. El valor numérico de esta constante, alrededor de 1/137, ha constituido desde entonces un tema incesante de especulación por parte de los físicos.

Para conseguir un tratamiento más satisfactorio del desdoblamiento de rayas hubo que esperar a los años 20, cuando se introdujo la mecánica cuántica. Se idearon dos formulaciones de la mecánica cuántica, una por Erwin Schrödinger y la otra por Werner Heisenberg y Max Born; pronto vieron que eran matemáticamente equivalentes. Un concepto esencial en ambas formulaciones es que el movimiento de un electrón no puede definirse de forma precisa, sino que debe describirse en términos de probabilidades. Ambas teorías predicen los niveles de energía de Bohr del átomo de hidrógeno, y, por consiguiente, las mismas longitudes de onda de las rayas espectrales; las predicciones pueden extenderse a sistemas atómicos más complicados. De nuevo deben tenerse en cuenta las correcciones relativistas para explicar el desdoblamiento de las rayas del hidrógeno.

Sin embargo, en la década de 1920-1930, se disponía ya de mejores medidas de la forma de las rayas en el espectro del hidrógeno, y pronto resultó evidente que eran necesarios nuevos reajustes de la teoría. El reajuste se hizo cuando se reconoció que el electrón podía adquirir momento angular no sólo por su giro alrededor del núcleo sino también por giro sobre su eje. El concepto de spin del electrón fue introducido por George Uhlenbeck y Samuel A. Goudsmit para explicar las medidas de los espectros de los átomos de los metales alcalinos. Al incorporar en la teoría el spin del electrón, se pudo predecir con exactitud razonable no sólo las longitudes de onda de las rayas espectrales sino también sus intensidades; lo cual no era cierto en la teoría de Bohr-Sommerfeld. En 1928, se fundamentó mejor el concepto de spin del electrón –y también del protón y del neutrón– cuando P. A. M. Dirac encontró una nueva forma de construir una teoría de la mecánica cuántica que era también coherente con la teoría de la relatividad especial. En la teoría de Dirac, el spin del electrón aparece como una consecuencia natural de las ecuaciones básicas y no como un postulado fundamental. La magnitud del momento angular de spin es $(1/2)(h/2\pi)$.



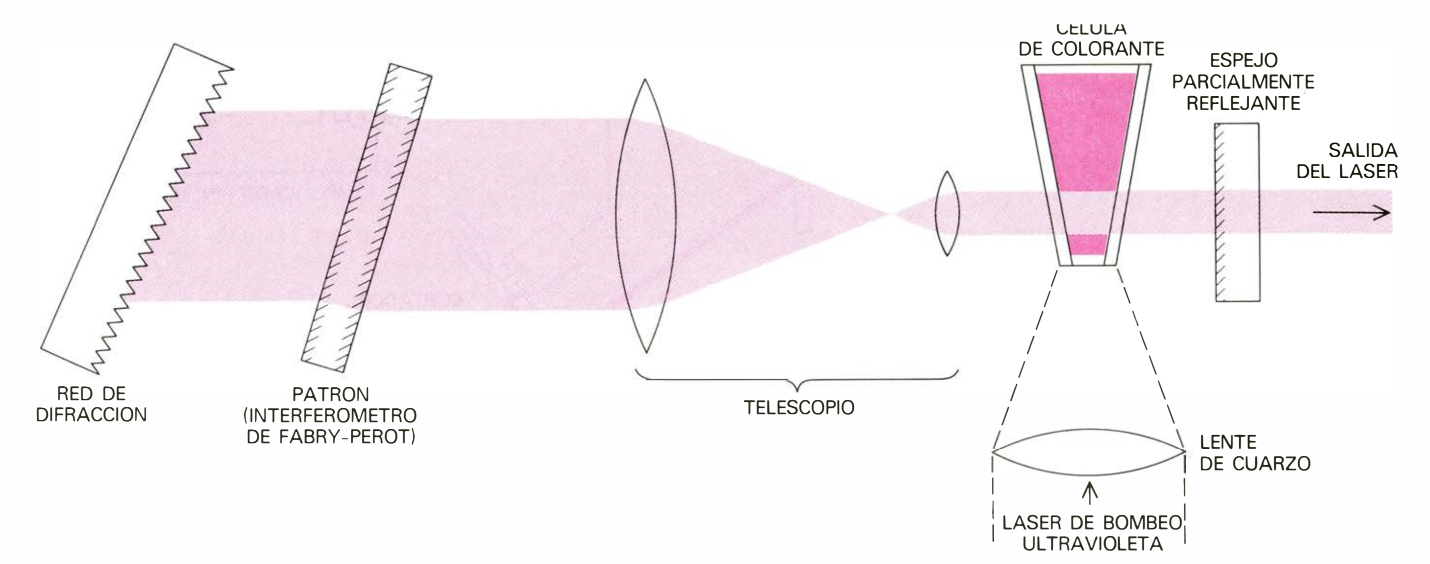
ENSANCHAMIENTO DOPPLER, que obscurece muchos de los detalles del espectro del hidrógeno, difuminando cada una de las estrechas componentes sobre una banda de longitudes de onda. Puesto que los átomos de un gas tienen velocidades distribuidas aleatoriamente, en cualquier instante algunos se están moviendo hacia una fuente de luz y otros se están alejando de ella. Para un átomo que se mueve hacia la fuente (*a*), la luz aparenta tener una frecuencia más alta (o longitud de onda más corta) que la que tiene en el sistema de referencia del laboratorio. Para un átomo que se aleje (*b*), la frecuencia disminuye. Como resultado de estos efectos Doppler, la frecuencia medida para una componente espectral es distinta para átomos que tengan velocidades diferentes. El perfil de absorción para todos los átomos del gas (*c*) no es un pico sino una curva ancha. Las técnicas de espectroscopia de láser eliminan el ensanchamiento Doppler.

Estructura fina

La mecánica cuántica relativista de Dirac predice la estructura del espectro del hidrógeno con gran detalle. Como en las teorías anteriores, cada uno de los niveles de energía básicos viene designado por un entero, n, que se llama número cuántico principal. Para el nivel de energía más baja (el estado fundamental del átomo), n es igual a 1; para el primer nivel excitado, n es igual a 2; y así sucesivamente. Estos enteros son los mismos que aparecen en la fórmula de Rydberg para las rayas espectrales.

Las rayas básicas del espectro –olvidando de momento cualquier desdoblamiento en componentes más finas– son generadas por transiciones entre estados que tienen distintos números cuánticos principales. La raya Balmer-alfa, por ejemplo, proviene de transiciones entre los estados $n = 2$ y $n = 3$. Si un átomo de hidrógeno cae de $n = 3$ a $n = 2$, emite luz de la longitud de onda de la raya Balmer-alfa; si el átomo absorbe entonces un cuanto de luz de la misma longitud de onda, realiza la transición opuesta, desde $n = 2$ a $n = 3$. Todas las rayas de una serie tienen común el estado de energía más baja. Las transiciones que dan origen a la serie de Balmer conectan estados con números cuánticos principales 2 y 3, 2 y 4, 2 y 5, y así sucesivamente. En la serie de Lyman, todas las transiciones empiezan en el estado con $n = 1$.

Las rayas básicas del espectro del hidrógeno reflejan la interacción dominante en el átomo de hidrógeno: la atracción electrostática entre el núcleo y el electrón. El desdoblamiento de cada raya refleja interacciones subsidiarias, efectos relativísticos y efectos magnéti-



LASER DE COLORANTE. Emite luz que puede ser confinada a un intervalo estrecho de longitudes de onda, pero también puede sintonizarse sobre una amplia banda. El medio en el láser es un colorante fluorescente que emite luz con un espectro ancho y esencialmente continuo cuando el colorante es "bombeado" por otro láser. De esta amplia banda de emisión, se selecciona un intervalo mucho más estrecho de longitudes de onda para amplificarlo en una cavidad óptica que está sintonizada para que resuene a una determinada frecuencia. El sintonizado aproximado se obtiene mediante una red de difracción, que dispersa la luz de forma que sólo determinadas longitudes de onda son reflejadas de nuevo hacia la célula de colorante. Un patrón limita aún más la anchura de banda dejando pasar sólo aquellas longitudes de onda que interfieren constructivamente durante muchas reflexiones internas. Un telescopio dentro de la cavidad esparce el haz de forma que queden iluminadas muchas rayas de la red. La luz monocromática pero sintonizable ha sido usada en medidas espectroscópicas que eliminan el ensanchamiento Doppler. El láser aporta luz de brillo, direccionalidad y gran pureza espectral.

cos. Los efectos magnéticos son el resultado de que el electrón, en razón de su spin, se comporta como un imán, que interacciona con el campo magnético atribuible al movimiento efectivo del núcleo cargado alrededor del electrón. Este desdoblamiento se conoce por estructura fina del espectro. Como el núcleo también tiene spin y actúa como un imán, interacciona con el campo magnético causado por el movimiento de la carga del electrón y también por la proximidad del campo magnético del electrón. Estas interacciones magnéticas son aún de menor importancia, y los desdoblamientos correspondientes de las rayas espectrales se conocen con el nombre de estructura hiperfina.

Se ha puesto a punto un sistema para clasificar los niveles de energía finos e hiperfinos, que especifica las magnitudes de los distintos momentos angulares (orbital, spin electrónico, spin nuclear y total) en un estado estacionario. El número cuántico principal, que corresponde al nivel de energía de Bohr, viene dado por uno de los enteros de la secuencia 1, 2, 3... A este número se le añade una letra, tal como *S, P, D* o *F*, que está relacionada con el momento angular orbital: las letras simbolizan los enteros de la serie que empieza por 0, 1, 2, 3... Un tercer símbolo, que para el átomo de hidrógeno es siempre una fracción tal como 1/2, 3/2 o 5/2, está relacionado con la suma del momento angular orbital y de spin. Si un análisis es suficientemente sutil para revelar la estructura hiperfina (una interacción que depende de las propiedades magnéticas del núcleo), debe añadirse una cuarta cantidad, representada por el símbolo *F*. Está relacionada con la suma de los momentos angulares orbital, de spin del electrón y de spin del núcleo. Generalmente, estados estacionarios que difieran en uno cualquiera de estos números tienen energías distintas, aun cuando éste no es siempre el caso.

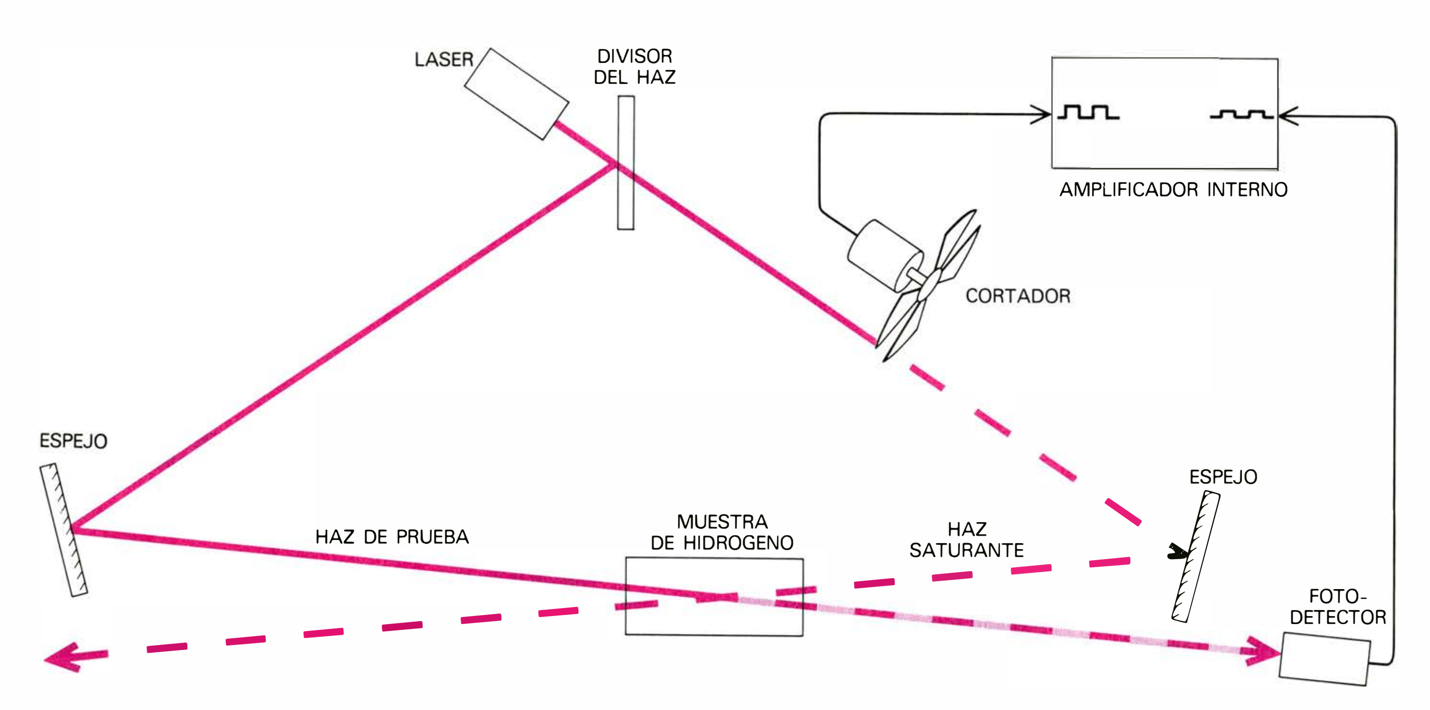
La estructura fina divide el estado $n = 2$ de un átomo de hidrógeno en tres subniveles y el estado $n = 3$ en 5 subniveles. Una simple tabla de las posibles combinaciones de estos niveles implica que la raya Balmer-alfa podría tener 15 componentes. En realidad, varias de estas transiciones están prohibidas por leyes de conservación, y la teoría de Dirac predice que la raya debería tener siete componentes. Cuando Dirac propuso su teoría, los experimentales podían reconocer claramente sólo dos componentes, aunque en 1925 G. Hansen detectó una tercera componente como un ensanchamiento lateral de una de las otras.

El factor que limita la resolución de las raya era (y sigue siendo) el ensanchamiento Doppler causado por los movimientos térmicos de los átomos. A causa del efecto Doppler, la luz emitida por un átomo aparece desplazada a frecuencias más altas si el átomo se aproxima al observador y a frecuencias más bajas si el átomo se aleja de éste. Los átomos de las fuentes luminosas ordinarias se mueven en todas las direcciones y, por consiguiente, la luz emitida se ensancha en longitud de onda en un factor v/c, donde v es la velocidad media de los átomos y c es la velocidad de la luz. El ensanchamiento es particularmente molesto para el hidrógeno, cuyos átomos son los más ligeros y, por tanto, se mueven rápidamente. A la temperatura ambiente, la anchura fraccional de raya es de 12 partes por millón. La velocidad de los átomos, y, por tanto, la anchura de la raya es proporcional a la raíz cuadrada de la temperatura absoluta. La anchura de raya puede ser reducida en un factor de tres o cuatro enfriando la lámpara de descarga de hidrógeno desde la temperatura ambiente (300 grados Kelvin) a la temperatura del hidrógeno líquido (20 grados K).

Efecto Lamb

En la década que se inicia en 1930 varios investigadores trataron de resolver la estructura fina del hidrógeno de forma más completa para comprobar las predicciones de la teoría de Dirac. Por aquel entonces, se podía examinar ya el espectro del deuterio, el isótopo del hidrógeno con una masa dos veces mayor que la del hidrógeno ordinario. La anchura de una raya espectral varía inversamente con la raíz cuadrada de la masa atómica y, por tanto, la anchura queda reducida en el espectro del deuterio en un factor 1,4. Aun así, no se encontraron nuevas componentes de la raya Balmer-alfa, y la tercera componente todavía no quedaba totalmente separada de su vecina más intensa.

En el transcurso de estas investigaciones algunos observadores hicieron notar la presencia de pequeñas discrepancias entre los espectros medidos y las predicciones de la teoría de Dirac. De interés especial eran el par de niveles de estruc-



ESPECTROSCOPIA DE SATURACION. Elimina el ensanchamiento Doppler seleccionando un grupo de átomos que carecen de componente del movimiento a lo largo del eje óptico. La luz de un láser de colorante se divide en un haz intenso saturante y un haz más debil de prueba, que pasan a través de una muestra de hidrógeno atómico en sentidos opuestos. El haz saturante es suficientemente intenso para poder reducir en forma notable la población de átomos capaces de absorber la longitud de onda del láser: el haz despeja un camino a través del gas. El haz de prueba encuentra, por tanto, una absorción más pequeña o registra una intensidad más alta en el detector. Sin embargo, los dos haces pueden interaccionar de esta forma sólo cuando ambos son absorbidos por los mismos átomos del gas; esto sucede cuando ambos están sintonizados a la longitud de onda de los átomos que no tienen desplazamiento Doppler. En la práctica, el haz saturante es interrumpido periódicamente con un dispositivo; se detecta un aumento en la transmisión del haz de prueba sintonizando el láser a través de un intervalo de longitudes de onda y buscando una señal a la frecuencia de interrupción.

a
HAZ INTENSO SATURANTE
HAZ DEBIL DE PRUEBA

b

c

HACES QUE SE PROPAGUEN en sentidos opuestos, procedentes de un láser sintonizable, sufrirán desplazamientos Doppler opuestos. Cuando la frecuencia del haz saturante (medida en el sistema de referencia del laboratorio) está por debajo de alguna componente seleccionada del espectro (*a*), sólo los átomos que se mueven hacia la izquierda pueden absorber luz, porque su movimiento desplaza la frecuencia para que resuene con la transición atómica. El haz de prueba, con la misma frecuencia, es absorbido sólo por átomos que se mueven hacia la derecha; por consiguiente, no se ve afectado por el haz saturante. Cuando la frecuencia del láser está por encima de la componente espectral (*b*), el haz saturante es absorbido por átomos que se mueven hacia la derecha y el haz de prueba por átomos que se mueven hacia la izquierda, que constituyen de nuevo dos grupos distintos de átomos. Sólo cuando la frecuencia del láser es igual a la frecuencia de la transición atómica (*c*), los dos haces interaccionan con los mismos átomos; los que están en reposo efectivo. El haz de prueba encuentra entonces que átomos que hubieran podido contribuir a la absorción son incapaces de hacerlo, pues ya han absorbido radiación del haz saturante. Sólo se representa la componente del movimiento que está según el eje óptico.

tura fina $2S_{1/2}$ y $2P_{1/2}$. Estos niveles difieren en el momento angular orbital y, por tanto, en la distribución geométrica media del electrón, pero tienen el mismo momento angular total, y la teoría de Dirac predice que deberían tener la misma energía. Por consiguiente, aunque son estados distintos, no deberían dar lugar a ningún desdoblamiento en el espectro. Simon Pasternack interpretó los experimentos llevados a cabo por Robley C. Williams y por William V. Houston y sus colegas como muestras de que los niveles $2S_{1/2}$ y $2P_{1/2}$ no coinciden exactamente. Pero había quienes no estaban de acuerdo.

El problema fue resuelto definitivamente en 1947 mediante un brillante experimento realizado por Willis E. Lamb, Jr., y Robert C. Retherford. Encontraron que el nivel $2S_{1/2}$ estaba desplazado hacia arriba con respecto al nivel $2P_{1/2}$ en unos 1060 megahertz (millones de ciclos por segundo). Comparado con la frecuencia de la raya Balmer-alfa, que es de unos 450.000 gigahertz (miles de millones de ciclos por segundo), esto es un desplazamiento en la posición de la componente $2S_{1/2}$ por un factor ligeramente superior a dos partes por millón.

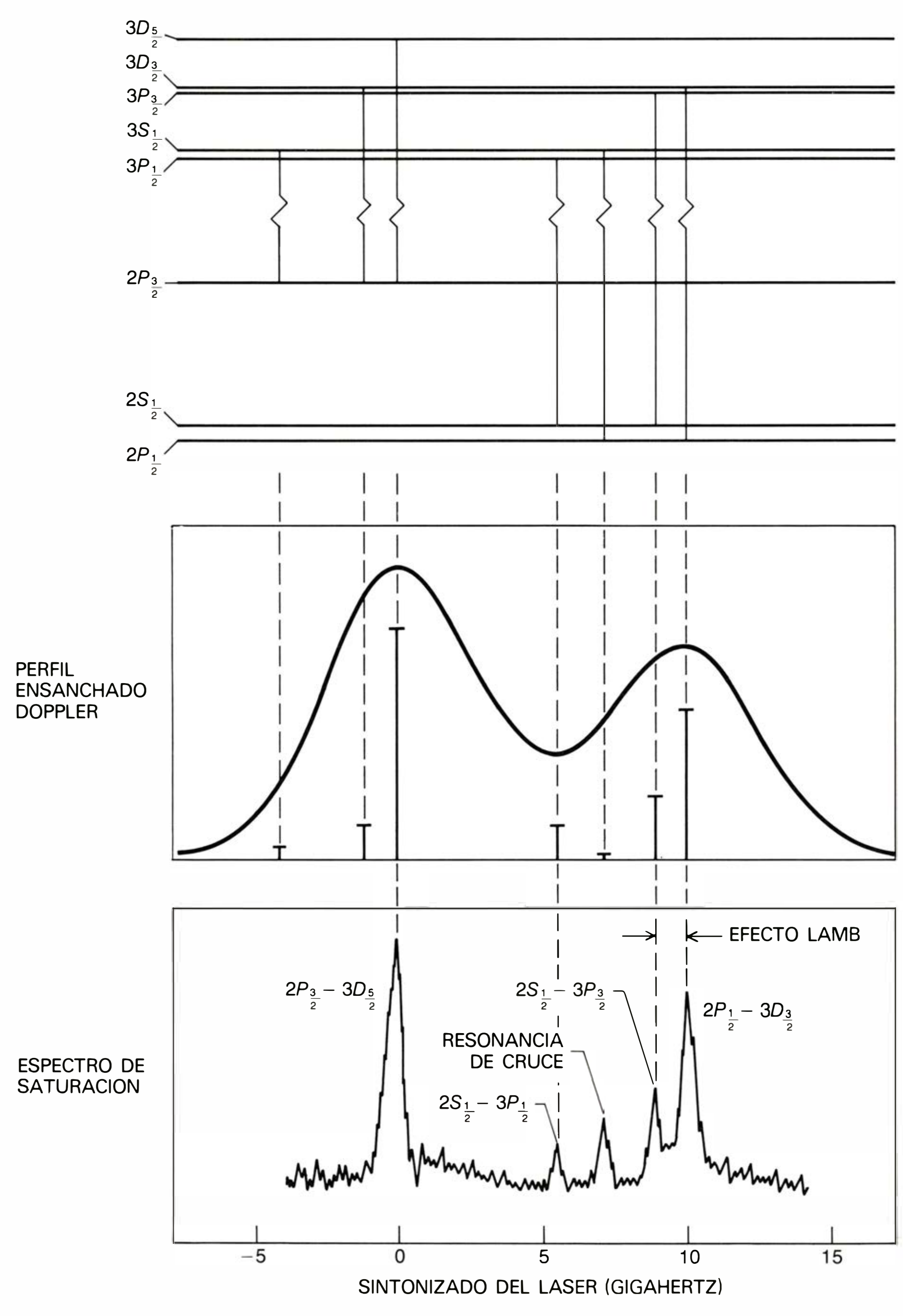
Lamb y Retherford no intentaron resolver las dos componentes en el espectro óptico. En su lugar, ellos emplearon energía de radiofrecuencia para estimular una transición directa entre los niveles $2P_{1/2}$ y $2S_{1/2}$. La frecuencia de las ondas de radio es 500.000 veces más baja que la frecuencia de las ondas Balmer-alfa; como el ensanchamiento Doppler de la línea queda reducido por el mismo factor, puede ser totalmente despreciado. A este pequeño desdoblamiento de los niveles energéticos se le llama hoy efecto Lamb.

La necesidad de dar una explicación del efecto Lamb inspiró otra revisión fundamental de la teoría física, introducida por Richard P. Feynman, Julian Schwinger y Sin-itiro Tomonaga. Su punto de partida era una nueva idea de que los electrones que se observaban en los experimentos podían diferir en ciertas propiedades, en masa y carga, por ejemplo, de los electrones hipotéticos "desnudos" que existirían en un universo vacío. Estas diferencias, que hasta entonces habían sido ignoradas, pueden alterar las propiedades de un electrón en un estado ligado. La teoría que ha surgido de estas ideas se llama electrodinámica cuántica.

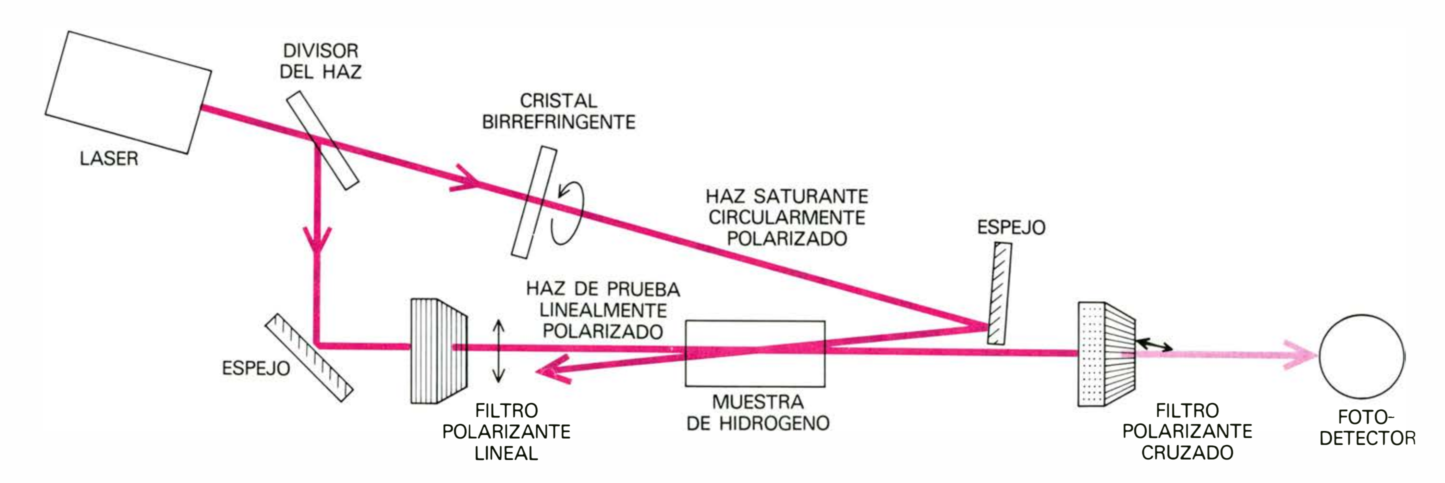
En electrodinámica cuántica, el efecto Lamb tiene su origen en la interacción entre el electrón y las "fluctuaciones de punto cero" del campo electromagnético. Tales fluctuaciones, que están siempre presentes, causan un temblor del electrón, de forma que percibe el núcleo algo difuminado en el espacio. El resultado es una alteración aparente de la fuerza entre las dos partículas cuando se hallan muy próximas. Como el electrón está más tiempo cerca del núcleo en el estado S que cuando se halla en el estado P, la energía del estado S queda ligeramente aumentada.

Este modelo del átomo de hidrógeno tiene una nueva complicación. Si el electrón se considera como una partícula puntual, puede responder a fluctuaciones de frecuencia infinitamente alta, que tienen energía infinita. Hay que evitar los infinitos; y así topamos con la necesidad de tener que responder a ciertas preguntas sobre el tamaño y la estructura del electrón, comparando un electrón ligado a un átomo con un electrón hipotético desnudo. La pequeña energía medida se encuentra como una diferencia. Estos cálculos son complicados, pero las predicciones de la teoría son extraordinariamente precisas. En efecto, la electrodinámica cuántica es la teoría más precisa jamás construida y no se han hallado discrepancias entre la teoría y los experimentos.

Se ha logrado poner de manifiesto,



LA LINEA BALMER-ALFA se resuelve en una serie de picos mediante espectroscopía de saturación. Los cálculos teóricos indican que la línea tiene siete componentes como consecuencia de los desdoblamientos causados por la estructura fina y el efecto Lamb, e ignorando la estructura hiperfina que es de un orden de magnitud mucho menor. En el perfil ensanchado por efecto Doppler, sólo se pueden distinguir dos componentes. Examinando únicamente aquellos átomos que no tienen efecto Doppler mediante espectroscopía de saturación aparecen cuatro componentes (un quinto pico a mitad de distancia de otros dos picos es un efecto espúreo, que recibe el nombre de resonancia de cruce). La escala de frecuencias mide la sintonización del láser desde un punto de partida que no es fijo, sino arbitrario.



ESPECTROSCOPIA DE POLARIZACION. Se basa en el hecho de que pequeños cambios en la polarización de la luz se detectan más fácilmente que pequeños cambios en la intensidad. De nuevo, la salida del láser se divide en dos haces, pero se da al haz saturante una polarización circular y al haz de prueba una polarización lineal. Como el haz de prueba encuentra filtros polarizantes cruzados, sólo una parte muy pequeña de él puede alcanzar el detector a menos que el gas de hidrógeno consiga hacer girar de alguna forma el plano de polarización del haz. Tal rotación puede ser producida por el haz saturante, que está circularmente polarizado, sólo si los dos haces interaccionan con la misma población de átomos. Los átomos que están estacionarios en un cierto instante son seleccionados por su posibilidad de interaccionar con ondas luminosas que se mueven en sentidos opuestos.

CAMBIO EN LA POLARIZACION de un haz de prueba, que es inducido por el paso anterior de un haz saturante. Como el haz saturante está circularmente polarizado es absorbido selectivamente por átomos que tienen una orientación particular (*a*); en consecuencia, estos átomos son eliminados de la población de átomos que pueden absorber radiación a la longitud de onda del láser. La mayor parte de los átomos restantes tienen la orientación opuesta. El haz de prueba linealmente polarizado puede considerarse como una combinación de dos ondas circularmente polarizadas que poseen igual intensidad, pero sentidos contrarios de rotación (*b*). Cuando el haz de prueba pasa a través de la muestra, una de las componentes circularmente polarizadas se absorbe de un modo más fuerte, puesto que los átomos están orientados preferentemente en una dirección. La absorción selectiva de una de las componentes circularmente polarizadas cambia la polarización de la luz restante (*c*) y, por tanto, una parte del haz de prueba puede atravesar el filtro analizador. El cambio de la polarización sólo puede ocurrir, sin embargo, si los dos haces son absorbidos por la misma población de átomos.

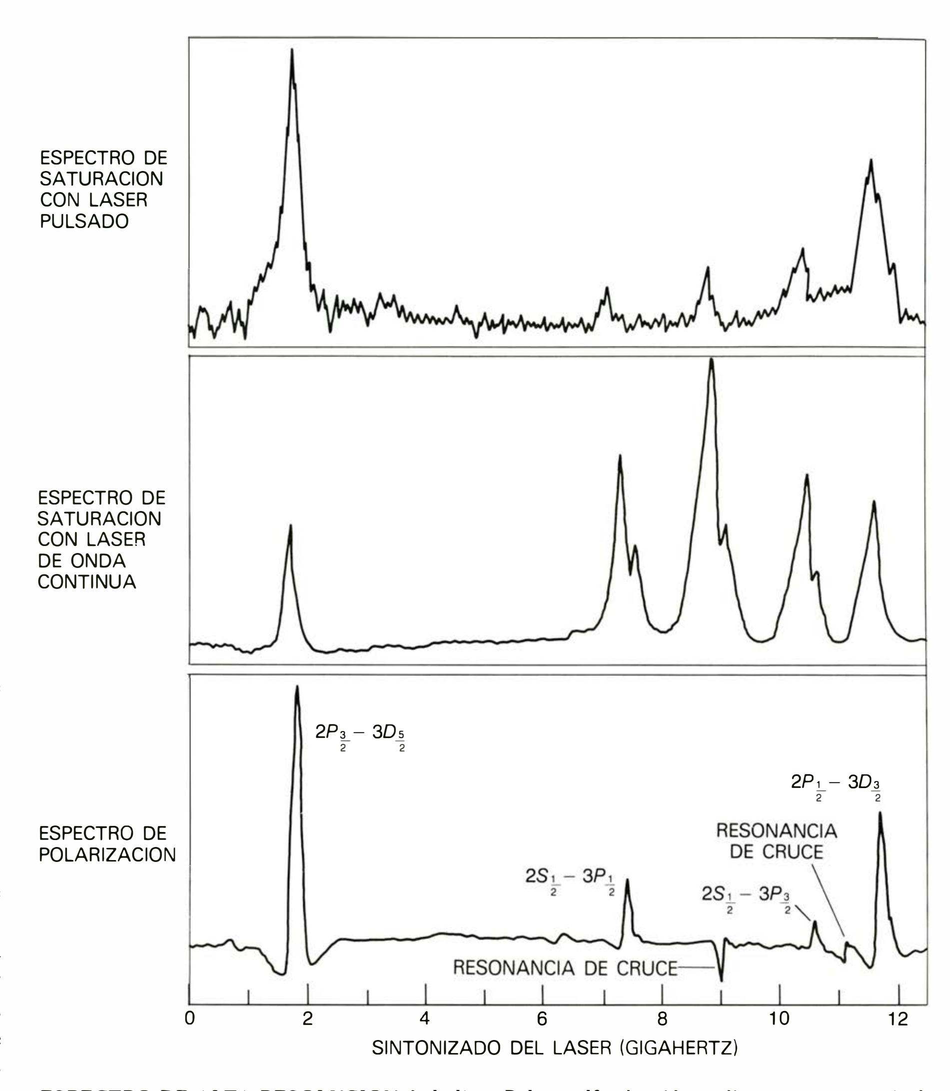
con notable detalle, el desdoblamiento debido a la estructura fina de diversos niveles del átomo de hidrógeno a través de mediciones de radiofrecuencia. Al mismo tiempo, métodos espectroscópicos mejorados han conducido a determinaciones precisas de la constante de Rydberg a partir de mediciones de la longitud de onda de una componente de la raya Balmer-alfa. Bryan P. Kibble, William R. C. Rowley, Raymond E. Shawyer y uno de nosotros (Series) emplearon una fuente luminosa enfriada en helio líquido para determinar la constante de Rydberg con una precisión más exacta que una parte en 10 millones. Incluso se observó el efecto Lamb en 1948 por uno de nosotros (Series) y fueron medidos desplazamientos análogos en helio ionizado por uno de nosotros (Series), por Gerhard Herzberg y por Frederick L. Roesler. Sin embargo, la precisión de los experimentos ópticos no se aproxima a los métodos de radiofrecuencias.

Anchura de las rayas espectrales

Si bien estos experimentos se han visto coronados por el éxito, el átomo de hidrógeno constituye un marco tan fundamental para comprobar las leyes de la física que es necesario estudiarlo mejor conforme se introducen nuevos métodos experimentales. Por ahora, el enfoque más prometedor es el de mejorar la resolución de los espectros medidos; la resolución limita la nitidez de las rayas espectrales observadas y, por tanto, la precisión con que se pueden determinar sus longitudes de onda.

Los primeros estudios espectroscópicos de los átomos venían condicionados por la resolución del mismo instrumento, que era simplemente un prisma o una red de difracción, una placa de cristal surcada por muchas rayas muy juntas. La red dispersa la luz al reflejar distintas longitudes de onda a distintos ángulos. Si todo el espectro visible es dispersado, digamos, en 15 grados, entonces algunas de las componentes de la raya Balmer-alfa son separadas menos de 0,05 segundos de arco.

Con la aplicación de los interferómetros al examen de las rayas espectrales la limitación intrínseca de los instrumentos desapareció. El interferómetro inventado por Michelson y otro tipo diseñado por los ópticos franceses Charles Fabry y Alfred Perot son instrumentos muy sensibles para comparar dos longitudes de onda. Pueden distinguir entre dos picos que difieran, en frecuencia, un gigahertz o menos. Los interferómetros fallan al intentar separar completamente las distintas componentes de las rayas



ESPECTRO DE ALTA RESOLUCION de la línea Balmer-alfa obtenido mediante espectroscopía de polarización. Para que se pueda comparar se repite también en la parte superior el espectro de saturación que se muestra en la página 41 y un espectro de saturación mejorado, hecho con un láser de onda continua en lugar de usar uno pulsante en la parte media. En el espectro de polarización no se pueden detectar componentes adicionales, pero los cuatro picos que se han resuelto aparecen más nítidos. Los dos picos pequeños entre los altos más externos (excluyendo resonancias de cruce) tienen una anchura media de alrededor de 35 megahertz, casi 10 veces más estrechos que las correspondientes rayas en el espectro de saturación pulsado. El espectro de polarización tiene una apariencia poco común (se extiende por debajo del nivel fundamental como también por encima de él), porque la gráfica corresponde al cambio en intensidad cuando el láser es escudriñado a lo largo de su intervalo de frecuencia.

Balmer, no porque los instrumentos no sean adecuados sino porque las componentes no son nítidas. Cada componente se distribuye sobre un intervalo de longitudes de onda que suele ser mayor que la distancia entre componentes.

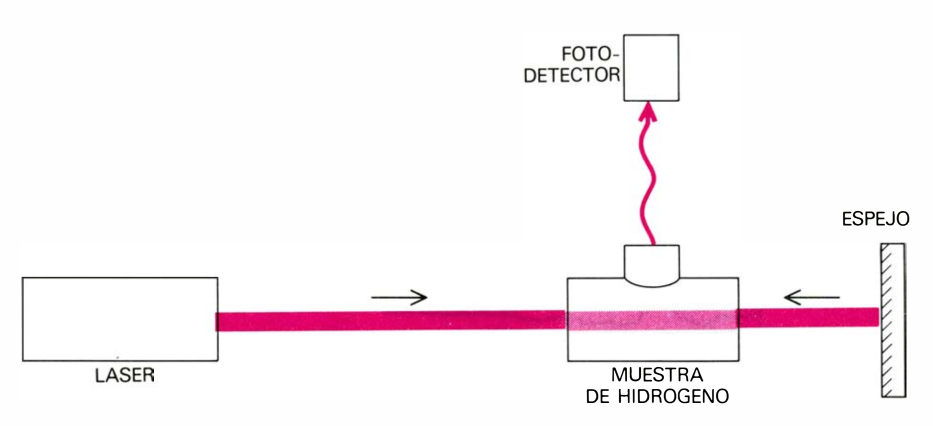
Una parte del ensanchamiento es la anchura "natural" de la raya espectral. Aún en las mejores circunstancias posibles de observación, ninguna componente del espectro puede ser absolutamente monocromática. La limitación obedece a que los "estados estacionarios" de un átomo no son realmente estacionarios. Es más, los átomos radian energía después de un intervalo finito de tiempo. Esto equivale a decir que los estados excitados acaban por desintegrarse.

La anchura natural de las rayas establece un límite final a la resolución espectroscópica, pero hasta hace poco nos habíamos acercado muy escasas veces a ese límite. Ello se debe a que las rayas son ensanchadas en mucha mayor extensión por otro efecto: el desplazamiento Doppler de las frecuencias como resultado del movimiento de los átomos.

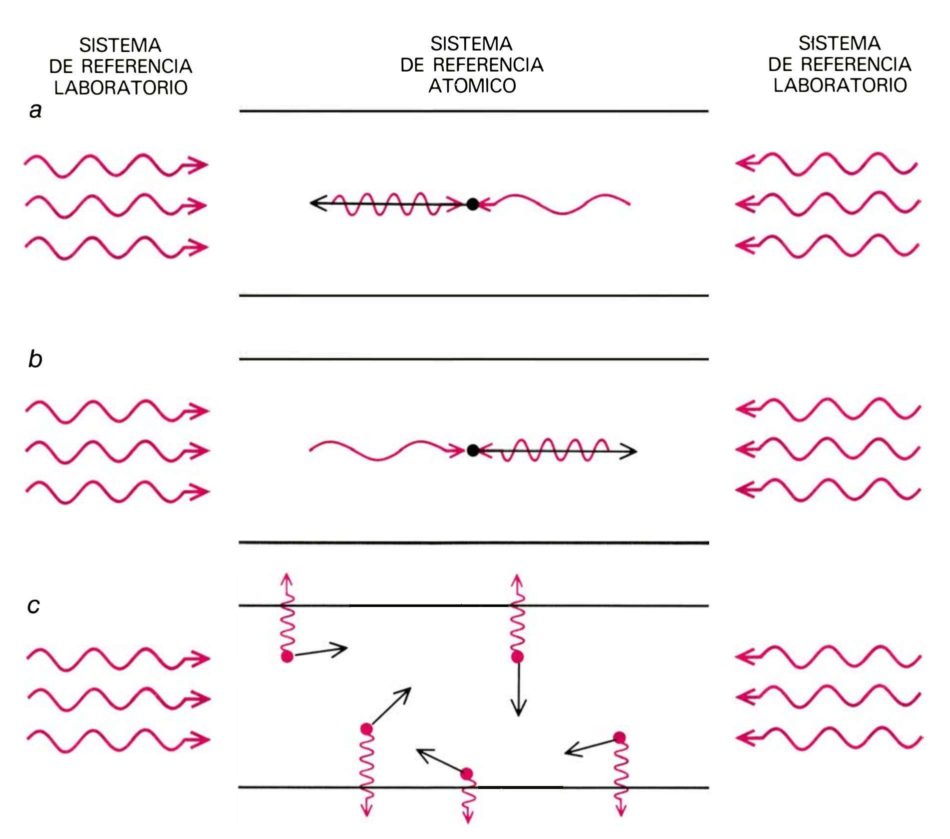
Para reducir notablemente el ensanchamiento Doppler, existe un método que consiste en medir el espectro, no de un gas, sino de un haz atómico colimado. Se puede crear un haz de este tipo dejando que los átomos de un gas escapen de una vasija a través de una serie de aperturas del tamaño de una aguja. Todos los átomos que puedan pasar por el agujero se moverán en la misma dirección y, por tanto, la luz que cruce el haz en ángulo recto con el haz tendrá unos desplazamientos Doppler pequeños. Sin embargo, el número de átomos

del haz irá disminuyendo al mejorar la colimación. Más aún, no siempre se puede preparar un haz de átomos en el estado deseado. Para estudiar el espectro Balmer del hidrógeno por absorción, por ejemplo, los átomos deben hallarse en el estado con $n = 2$; con tales átomos, no se pueden construir haces suficientemente densos para las medidas espectroscópicas convencionales.

Las técnicas de espectroscopía de láser introducidas desde aproximadamente el año 1970 han mejorado la resolución de forma notable porque incluyen métodos de eliminar el ensanchamiento Doppler. Tres métodos de este tipo serán discutidos aquí. Dos de ellos funcionan seleccionando para la observación únicamente aquellos átomos del gas que tienen componente nula del movimiento a lo largo del eje óptico. El tercer método deja que todos los átomos en un gas contribuyan a una medición, pero elimina el ensanchamiento debido a la velocidad disponiendo dos desplazamientos Doppler en sentidos opuestos.



LA ESPECTROSCOPIA DE DOS FOTONES no selecciona los átomos sin efecto Doppler sino que cancela los desplazamientos Doppler de todos los átomos en el gas. El láser de colorante es sintonizado a una frecuencia que es exactamente la mitad de la transición atómica seleccionada. El haz pasa a través de la muestra y después se refleja sobre sí mismo, creando un campo de ondas estacionarias. Los átomos del gas pueden realizar la transición seleccionada absorbiendo a la vez dos fotones, o cuantos de luz, que vienen de dos direcciones opuestas y, por tanto, tienen desplazamientos Doppler opuestos. Los estados excitados que se crean así se desintegran a algún nivel de energía más bajo, emitiendo fotones.



LOS DESPLAZAMIENTOS DOPPLER OPUESTOS se cancelan en la espectroscopía de dos fotones. Un átomo que se mueve hacia la izquierda (*a*) ve la onda que proviene de la izquierda desplazada a una frecuencia más elevada, pero la frecuencia de la onda que proviene de la derecha es desplazada hacia abajo en la misma cantidad. Para un átomo que se mueve hacia la derecha (*b*), los desplazamientos de las frecuencias son exactamente los opuestos. Un átomo estacionario no percibe desplazamientos Doppler. Independientemente de la velocidad del átomo, la suma de las dos frecuencias es constante e igual a dos veces la frecuencia del láser (medida en el sistema de referencia del laboratorio). Todos los átomos pueden, por consiguiente, absorber dos fotones (*c*) y después radiar a una frecuencia más alta.

El láser sintonizable de colorante

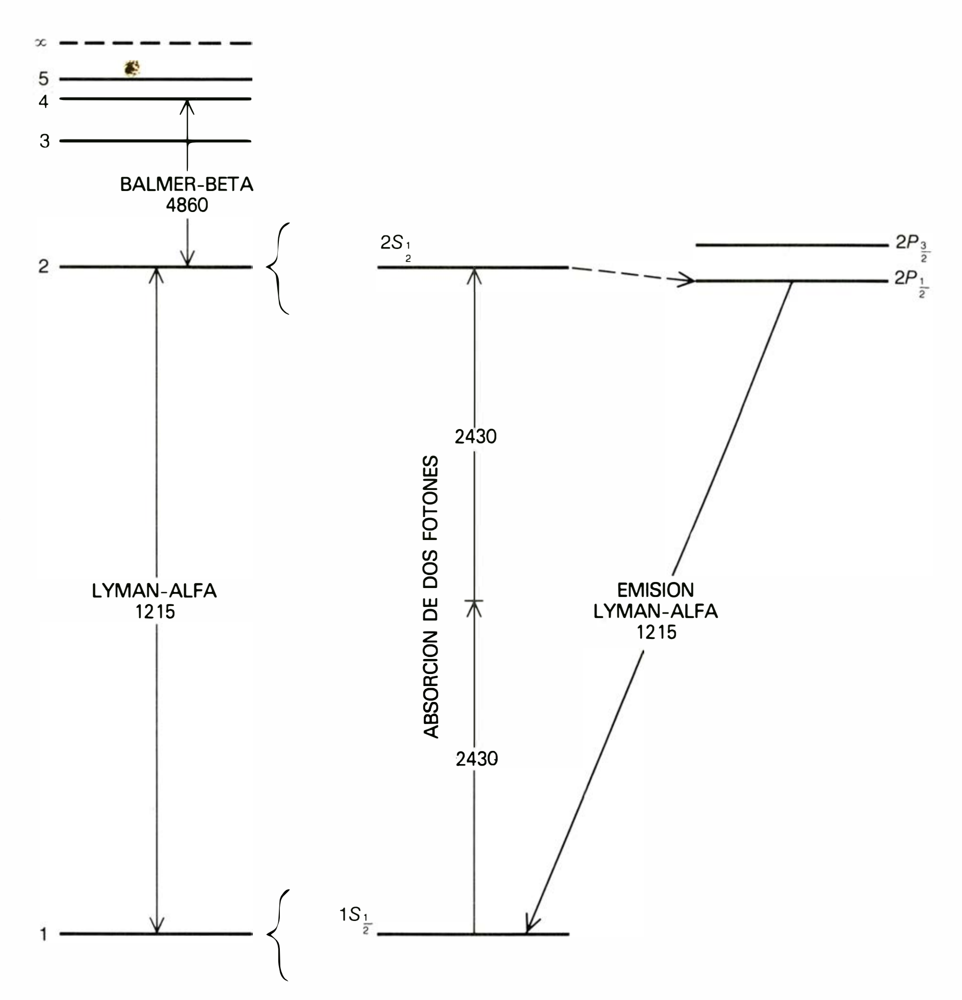
Los láseres nos dan luz de brillo, direccionalidad y pureza espectral no igualados. Sin embargo, los primeros láseres podían emitir luz sólo en unas pocas longitudes de onda discretas, determinadas por las rayas espectrales de los átomos o iones activos. Tal instrumento se podía usar en espectroscopía sólo si la transición que se quería observar coincidía con una de las rayas de láser disponibles.

En 1966, Peter Sorokin, del Centro de Investigaciones Thomas J. Watson de la International Business Machines Corporation, e, independientemente, Fritz Schäfer, de la Universidad de Marburg, descubrieron que disoluciones líquidas de ciertas moléculas orgánicas fluorescentes podían servir de medio para un láser con una banda ancha y continua de longitudes de onda. Estas moléculas orgánicas son colorantes que, por naturaleza, tienen bandas de absorción intensas; por ello están intensamente coloreadas. En el láser, las moléculas del colorante son "bombeadas", o llevadas a un estado excitado, por una fuente externa de luz intensa, otro láser por ejemplo.

Si bien el láser de colorante puede amplificar luz en un ancho intervalo de longitudes de onda, se encontró pronto que podía seleccionarse una determinada longitud de onda diseñando un resonador óptico que permitiera que sólo la longitud de onda elegida pasara repetidamente por la célula de colorante amplificadora. En 1970, uno de nosotros (Hänsch) ideó un sencillo láser de colorante que emitía pulsos de alta pureza espectral. El colorante, encerrado en una celda de vidrio, era bombeado mediante pulsos de radiación ultravioleta de un láser de nitrógeno. El resonador constaba, en un extremo, de una placa de vidrio parcialmente reflectante y, en el otro, de una red de difracción. La red dispersa la luz emitida por el colorante, de forma que sólo se refleja un intervalo estrecho de longitudes de onda a lo largo del eje óptico. Un telescopio instalado en la cavidad del láser expande el haz antes de que llegue a la red, de suerte que se iluminen muchos surcos con luz bien colimada. La anchura de raya del láser se reduce aún más si insertamos un interferómetro de Fabry-Perot inclinado dentro de la cavidad. El interferómetro

consta de dos superficies parcialmente reflectantes, cuya separación se controla de forma muy precisa; el tipo de interferómetro colocado en el resonador del láser se llama patrón y consiste en una placa de cristal cuyas dos superficies paralelas están revestidas con capas parcialmente reflectantes. Ondas luminosas con longitudes de onda distintas toman caminos ligeramente distintos en el patrón. Como resultado de múltiples reflexiones internas, algunas longitudes de onda interfieren destructivamente y se extinguen, mientras que otras longitudes de onda interfieren constructivamente y quedan reforzadas.

La anchura de raya de este láser es de unas pocas milésimas de angstrom (alrededor de una parte por millón), comparable a la resolución de los mayores espectrógrafos de red. La anchura puede reducirse todavía en un factor de 100, aproximadamente, colocando un segundo interferómetro de Fabry-Perot fuera de la cavidad del láser, el cual actúa como un filtro pasivo. Aunque, para cada posición, la luz es altamente monocromática, el láser puede sintonizarse sobre todo el intervalo del perfil de absorción del colorante. Puede lograrse un ajuste aproximado cambiando el ángulo de la red de difracción o del patrón. Se consiguen ajustes más finos encerrando la red y el patrón en una cámara y alterando la presión, lo cual cambia el índice de refracción del gas encerrado.



ESTADO FUNDAMENTAL DEL HIDROGENO. El estado fundamental, o más bajo, del hidrógeno puede ser alcanzado sólo mediante transiciones cuyas longitudes de onda están en la región ultravioleta del espectro electromagnético; tales transiciones pueden ser observadas más convenientemente mediante espectroscopia de dos fotones que por cualquier otra técnica. La línea Lyman-alfa, una transición de $n = 1$ a $n = 2$, tiene una longitud de onda de 1215 angstrom, y es necesaria una radiación de dos veces esta longitud de onda, 2430 angstrom, para la espectroscopia de dos fotones. Aún esta última longitud de onda es inaccesible a los láseres de colorante y debe ser generada doblando la frecuencia (o dividiendo por dos la longitud de onda) de un láser sintonizado a 4860 angstrom. El multiplicar por dos la frecuencia se logra mediante un cristal que bajo iluminación intensa emite el segundo armónico de la frecuencia incidente. La longitud de onda original del láser (4860 angstrom) es muy aproximadamente la longitud de onda de la raya Balmer-beta. La absorción de dos fotones en hidrógeno atómico estimula únicamente la transición $1S_{1/2}$ a $2S_{1/2}$. El átomo excitado pasa al estado $2P_{1/2}$ debido a los choques con los otros átomos, y de este estado vuelve al estado fundamental emitiendo un fotón de 1215 angstrom.

Espectroscopía de saturación

Hay un límite final en la resolución de longitud de onda para cualquier láser pulsado. La precisión con que se puede conocer la longitud de onda viene determinada por el número de ciclos que se dispone para su medición y, por tanto, por la longitud del pulso. Líneas más estrechas exigen un láser de onda continua. En 1970, Otis G. Peterson, Sam. A. Tuccio y Benjamin B. Snavely, de la Eastman Kodak Research Laboratories, presentaron el primer láser sintonizable de colorante de onda continua. Los principios de funcionamiento eran los mismos que los de un láser sintonizable de colorante pulsado, si bien la disolución del colorante se hacía fluir rápidamente para evitar el sobrecalentamiento. En el comercio existen ya láseres de colorante de onda continua muy perfeccionados. En varios laboratorios se han obtenido anchuras de raya muy pequeñas, del orden de unas pocas partes en 10^{12}, mediante estabilización electrónica de la frecuencia del láser.

Dada una fuente luminosa que sea monocromática y sintonizable, se puede medir un espectro de absorción haciendo pasar la luz a través de una muestra del gas y registrando de forma continua las frecuencias en el entorno de una raya del espectro. Para medir la intensidad de la luz transmitida nos serviríamos de un detector fotoeléctrico, que debería registrar un mínimo en la longitud de onda de cada componente de la raya. Sin embargo, con este esquema mediríamos sólo el perfil del ensanchamiento Doppler de la línea. Para aprovechar totalmente la gran posibilidad de resolución de un láser deben encontrarse medios para eliminar los desplazamientos Doppler.

A uno de estos métodos se le conoce por espectroscopia de saturación. Su funcionamiento depende de dos propiedades de la luz láser: la pequeña anchura de raya y la alta intensidad. Cuando un átomo absorbe un cuanto de luz, se excita a un nivel energético más alto y, por tanto, se le elimina temporalmente, de la población de átomos absorbentes. En un haz intenso de luz, los átomos de un gas pueden absorber cuantos a una velocidad mayor que a la que ellos pueden volver al estado original. Como resultado, la población de los átomos absorbentes queda disminuida de forma sustancial. La luz despeja un camino en el gas, y un segundo haz de la misma longitud de onda encuentra reducido el poder absorbente del gas.

Lamb fue el primero en darse cuenta de la importancia de este efecto en espectroscopía de alto poder resolutivo. Hizo notar que las dos ondas que viajaran en direcciones opuestas dentro de un láser podrían colaborar para saturar la absorción de una clase particular de átomos en el medio activo del láser. Suponiendo que la anchura de raya de la luz sea suficientemente estrecha, debe

interaccionar sólo con aquellos átomos cuyo movimiento desplace la longitud de onda hasta que resuene con una componente del espectro atómico. Si la frecuencia de la luz es más baja que la frecuencia de la componente de la raya, por ejemplo, entonces las dos ondas en la cavidad interaccionarán con átomos que se muevan hacia ellas, y serán átomos distintos para los dos haces que se mueven en sentidos opuestos. Hay una sintonización del láser, sin embargo, para la que ambos haces interaccionan con los mismos átomos. La sintonización que corresponde a la longitud exacta de la componente espectral, siendo los átomos que pueden absorber la luz aquellos que están en reposo o que se mueven transversalmente al eje óptico. Cuando el láser se sintoniza a esta longitud de onda, la población de átomos en el estado absorbente disminuye y la potencia de salida del láser se hace más pequeña. Este "pozo de Lamb" fue observado por primera vez por Ross A. McFarlane, William R. Bennett, Jr., y Lamb. Fue empleado para espectroscopía de alta resolución por Abraham Szöke y Ali Javan, del Instituto de Tecnología de Massachusetts, pero se demostró que era de utilidad limitada. Sólo pueden estudiarse las mismas transiciones láser, o bien aquellas líneas moleculares que coinciden con las longitudes de onda del láser gaseoso, y cualquier gas que se desee estudiar debe colocarse dentro del delicado resonador de un láser de onda continua.

En 1970, Christian Bordé, de la Universidad de París, y por vía independiente uno de nosotros (Hänsch) idearon una técnica para medir espectros de saturación fuera del resonador del láser. La luz de un láser sintonizable se divide en un haz saturante intenso y un haz de prueba más débil. Luego, se induce que los dos haces sigan el mismo camino a través de la celda gaseosa, pero en sentidos opuestos. Si el láser se desintoniza ligeramente de la frecuencia de una transición atómica, entonces los dos haces interaccionan con átomos distintos y el haz saturante no produce ningún efecto en la absorción del haz de prueba. Sin embargo, cuando el láser está sintonizado de forma que es absorbido por átomos estacionarios, el haz saturante tiene el efecto de disminuir la población de átomos en el estado seleccionado y el haz de prueba sufre una absorción más pequeña. En efecto, el haz saturante marca todos los átomos que tienen una componente cero del movimiento a lo largo del eje óptico e ignora todos los demás.

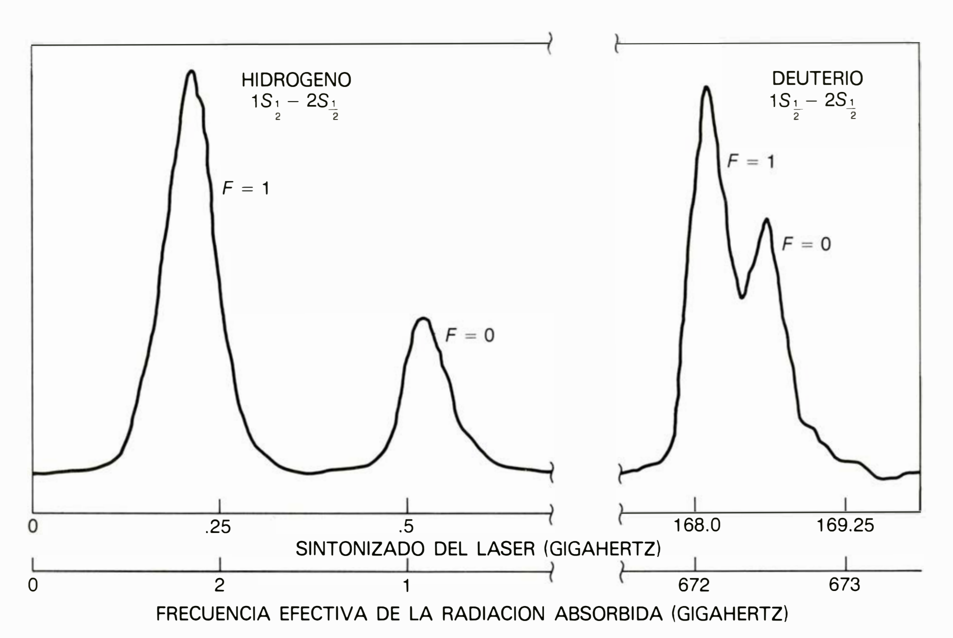
En la práctica, el cambio en absorción es pequeño y queda fácilmente obscurecido por el ruido. Para poder distinguir la señal del fondo, el haz saturante es cortado, o interrumpido a una frecuencia conocida, mientras el haz de prueba se deja que opere siempre en forma continua. Puede entonces detectarse la señal examinando la amplitud del haz de prueba para cualquier modulación a la frecuencia de corte.

Dos de nosotros (Hänsch y Schawlow) con Issa S. Shahin, que era entonces estudiante de tercer ciclo de la Universidad de Stanford, examinaron por este método la raya Balmer-alfa. Los átomos de hidrógeno eran excitados al nivel absorbente $n = 2$ en un tubo de descarga gaseosa de baja presión. Los dos haces luminosos de un láser de colorante pulsante pasaban a través de una sección del plasma de la descarga de una longitud de unos 15 centímetros, donde sufrían una absorción de alrededor de un 50 por ciento.

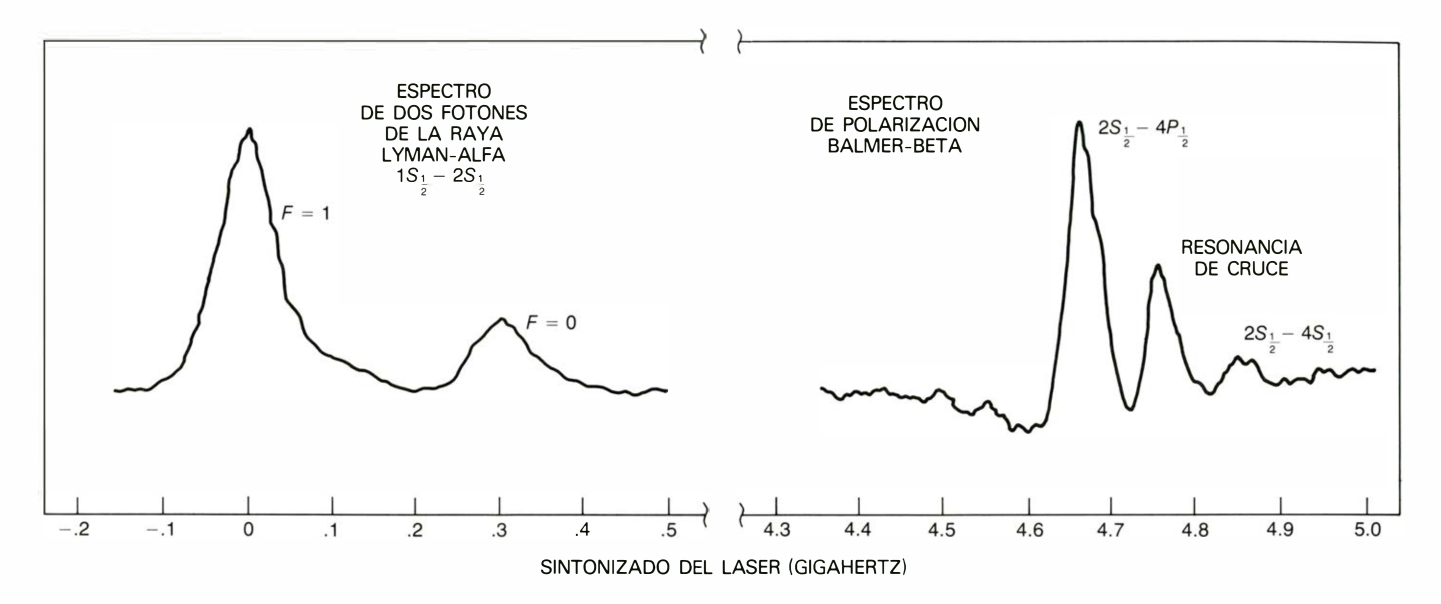
De las siete componentes de la estructura fina que se esperaban teóricamente, eran visibles cuatro; las dos más fuertes quedaban resueltas de forma mucho más nítida que en cualquier espectro anterior. Más aún, el efecto Lamb que separa los niveles $2S_{1/2}$ y $2P_{1/2}$ fue observado por primera vez en un espectro de absorción óptico.

El espectro de saturación incluía también un quinto pico, pero éste era espúreo, llamado resonancia de cruce, que aparece en el punto medio de dos transiciones que comparten el mismo nivel superior o inferior. Cuando el láser se sintoniza a la frecuencia correspondiente al punto medio de estas dos rayas, los átomos que se mueven en una dirección puede absorber el haz saturante, y los átomos que se mueven en sentido contrario pueden absorber el haz de prueba. En consecuencia, la absorción es saturada no por los átomos estacionarios sino por dos clases de átomos en movimiento.

La nitidez de los picos en el espectro de saturación sugirió que la constante de Rydberg podría determinarse con mejor exactitud midiendo la longitud de onda absoluta de una componente de la estructura fina. Esta medida fue abordada por un grupo de investigadores de Stanford dirigido por uno de nosotros (Hänsch) e incluyendo en particular a Munir Nayfeh. La componente elegida fue la transición $2P_{3/2}$ a $3D_{5/2}$, que es la más intensa, la que tiene el más pequeño desdoblamiento hiperfino no resuelto y la que se halla menos perturbada por el campo eléctrico del tubo de descarga. En primer lugar hubo que examinar los efectos sobre la posición de la raya debidos a las variaciones de la presión del gas, a la corriente y al voltaje en el tubo de descarga y a la intensidad del láser, y corregir posibles errores sistemáticos. Para medir la posición absoluta de la raya y no sólo el intervalo entre dos rayas, se necesitaba un patrón de longitud muy preciso. El patrón elegido fue un láser de helio-neón cuya frecuencia se bloqueaba electrónicamente hasta una particular componente hiperfina de una línea de absorción de vapor de iodo molecular; esta raya se definía también mediante espectroscopía de saturación. La longitud de onda del láser de he-



ESTRUCTURA HIPERFINA de la transición $1S_{1/2}$ a $2S_{1/2}$ que fue resuelta mediante espectroscopía de dos fotones, tanto en hidrógeno normal como en deuterio. Las transiciones entre los dos niveles hiperfinos en el hidrógeno son los responsables de la radiación de longitud de onda de 21 centímetros observada por los radioastrónomos. La anchura natural de las componentes es probablemente muy pequeña, quizá de un hertz; se resuelven dentro de unos 100 megahertz. El desplazamiento entre el hidrógeno y el deuterio (un isótopo suyo) se debe principalmente a la diferencia de masa de los núcleos.



MEDICION SIMULTANEA de la transición $1S_{1/2}$ a $2S_{1/2}$ y de la raya Balmer-beta, empleada para determinar la magnitud del efecto Lamb para el estado $1S_{1/2}$ del hidrógeno. La raya Balmer-beta fue medida por espectroscopía de polarización a la longitud de onda fundamental del láser, 4860 angstrom. Simultáneamente se usó la salida del mismo láser, doblada en frecu encia, para registrar el espectro de dos fotones de la transición $1S_{1/2}$ a $2S_{1/2}$ para longitud de onda efectiva de 1215 angstrom. Si los niveles de energía de Bohr fueran exactos, se observarían las dos transiciones a la misma sintonización del láser, pero los desdoblamientos de los niveles energéticos introducen pequeñas discrepancias. A partir de la diferencia medida en las sintonizaciones del láser para las dos rayas, se calculó la magnitud del efecto Lamb para el nivel $1S_{1/2}$ con una precisión 50 veces mayor que la mejor medición anterior. El valor encontrado para el hidrógeno es de 8,161 ± 29 megahertz. Los fecundos estudios sobre el átomo de hidrógeno no han agotado sus posibilidades.

lio-neón estabilizado era conocida con precisión en términos del patrón internacional de longitud, que es una raya espectral del kriptón emitida bajo condiciones especificadas. Los datos fueron evaluados en 1974. El resultado, $R = 10.973.731{,}43 \pm 0{,}10$ con dimensiones de recíproco de metros, era más preciso que el mejor valor anterior en un factor de casi 10.

Un método relacionado de espectroscopía libre de efecto Doppler fue introducido en 1976 por Carl Wieman, estudiante de tercer ciclo de Stanford que trabajaba con uno de nosotros (Hänsch). Aprovechando que los cambios pequeños en la polarización de la luz pueden detectarse más fácilmente que los cambios en la intensidad, esta técnica alcanza mucha mayor sensibilidad. Pueden observarse un menor número de átomos a intensidades más bajas de láser, con lo cual se evitan los desplazamientos y ensanchamientos sistemáticos de las rayas.

Espectroscopía de polarización

Como en la espectroscopía de saturación, la luz ael láser se divide en dos haces, uno más intenso que el otro, los cuales atraviesan la muestra en sentidos opuestos. En este caso, el haz más débil, o de prueba, ve la muestra de gas colocada entre filtros de polarización cruzados, de forma que la luz no alcanza prácticamente el detector mientras la muestra no cambie la polarización de la prueba. El haz saturante interacciona con la muestra para producir estos cambios. Puede hacerlo porque anteriormente se ha hecho pasar, a través de una lámina de cuarto de onda, una pieza de material birrefringente cortado y pulimentado de suerte que dé a la luz una polarización circular. Se dice que la luz está polarizada circularmente cuando el campo eléctrico gira (en el sentido de las agujas del reloj o en sentido contrario) en lugar de oscilar en un plano, como en el caso de luz linealmente polarizada.

La probabilidad de que un átomo absorba luz polarizada circularmente depende de la orientación del momento angular del átomo con relación al campo que gira. Al principio, la orientación del átomo es arbitraria, pero el haz saturante hace disminuir el número de átomos del gas que por su orientación sean capaces de absorber luz de una determinada polarización circular, dejando un exceso de átomos con el sentido opuesto de rotación. Cuando el haz de prueba, que está linealmente polarizado, pasa a través de la misma región del gas, los átomos orientados pueden alterar su propagación.

Una onda linealmente polarizada puede considerarse como la superposición de dos ondas circularmente polarizadas de igual intensidad, una que gire en el sentido de la agujas del reloj y la otra en sentido contrario. Cuando este haz pasa a través del gas, encuentra átomos que tienden a absorber sólo una de sus componentes circularmente polarizadas, atenuando ésta y dejando la otra componente más intensa. Debido a lo cual, cuando el haz de prueba sale del gas, está polarizado elípticamente y el eje de la elipse ha girado en relación con el plano de la polarización lineal inicial. El haz ha adquirido una componente que puede pasar a través del filtro de polarización cruzado. Todo esto puede suceder, sin embargo, sólo si el haz saturante y el haz de prueba interaccionan con los mismos átomos, esto es, aquellos que no tienen desplazamientos Doppler.

No hay luz transmitida por el filtro polarizante cruzado si los dos haces interaccionan con distintos conjuntos de átomos; por tanto, no hay esencialmente señal de fondo. En consecuencia, los picos reales del espectro no quedan fácilmente obscurecidos por el ruido o por las fluctuaciones de intensidad en el láser. Cuando la señal es muy pequeña suele resultar ventajoso girar ligeramente los filtros polarizantes. El detector registra entonces una señal de fondo determinada, que puede aumentar o disminuir, según la dirección en que se gira el plano de polarización.

John Goldsmith, estudiante de tercer ciclo de Stanford, Erhard Weber, que se encontraba en Stanford con permiso de la Universidad de Heidelberg y uno de nosotros (Hänsch) han obtenido recientemente un espectro mejorado de la línea Balmer-alfa mediante espectroscopía de polarización con láser. El haz saturante y el de prueba eran generados mediante un láser de colorante de onda continua y de baja potencia con una anchura de banda de menos de un megahertz.

El espectro se registraba en forma de una derivada, una función matemática que mide el ritmo de variación en la in-

tensidad del haz de prueba cuando el láser se sintonizaba a lo largo de una banda de frecuencias. En el espectro de polarización, los dos picos mayores son más nítidos todavía que en las mediciones anteriores por espectroscopía de saturación, y tres componentes menores resultan aún más estrechas. La anchura de raya para estas componentes es de unos 35 megahertz, casi diez veces menor que en las mediciones anteriores. Estas tres componentes representan todas transiciones que empiezan en el estado $2S_{1/2}$, que tiene una vida media larga; se sabe que cada uno de ellos se divide en dos componentes hiperfinas separadas por unos 178 megahertz. Parecería que el desdoblamiento hiperfino debería ser resuelto con claridad, pero sólo se observan picos únicos. La explicación de esta discrepancia es que, en cada caso, uno de los estados hiperfinos tiene momento angular total cero y no puede ser polarizado. Por tanto, cada pico observado no es una mezcla no resuelta de las dos transiciones sino una sola de las componentes hiperfinas.

La componente $2S_{1/2}$ a $2P_{1/2}$ fue elegida para hacer una nueva medida absoluta de la longitud de onda, calibrada de nuevo con la línea de referencia del iodo como longitud estándar. El nuevo valor de la constante de Rydberg obtenida mediante esta medida es $R = 10.973.731{,}476 \pm 0{,}032$ con dimensiones de recíproco de metros. La medida es casi tres veces más exacta que la anterior, y coloca a la constante de Rydberg entre las constantes fundamentales conocidas con más precisión. Sólo se han determinado con una precisión comparable la velocidad de la luz y el cociente de los momentos magnéticos del electrón y del protón.

Espectroscopía de dos-fotones

Estos métodos son útiles para las rayas de la serie de Balmer, es decir, aquellas rayas generadas por transiciones entre el nivel de energía $n = 2$ y niveles más altos. Para otros estados, la aplicabilidad del método no resulta tan fácil. Entre estos últimos es particularmente importante el estado $S_{1/2}$ con $n = 1$, que es el estado fundamental del átomo de hidrógeno. Incluso el grupo de estados más próximos, los estados con $n = 2$, están tan por encima del estado fundamental que la longitud de onda de la raya correspondiente, que se denomina Lyman-alfa, cae en la región ultravioleta del espectro electromagnético. La espectroscopía de precisión es muy difícil a estas longitudes de onda tan cortas, que se hallan mucho más allá del intervalo de operación del láser de colorante.

Sin embargo, con láseres intensos las dificultades de la espectroscopía del ultravioleta pueden evitarse excitando cada átomo mediante dos fotones, que juntos suministran la energía de excitación necesaria. La probabilidad de que un átomo absorba simultáneamente dos fotones es en general muy pequeña, pero crece con el cuadrado de la intensidad luminosa. La espectroscopía de dos fotones resulta de particular interés en el hidrógeno porque el ensanchamiento Doppler puede evitarse por un método ingenioso sugerido por primera vez en 1970 por L. S. Vasilenko, V. P. Chebotayev y A. V. Shishaev, del Instituto de Termofísica de Novosibirsk en la Unión Soviética. La puesta en marcha del método fue realizada en 1974 por tres grupos de investigadores que trabajaban independientemente. En lugar de dividir el haz del láser en dos componentes se refleja el haz sobre sí mismo por medio de un espejo, de forma que las ondas que se mueven hacia la derecha y hacia la izquierda viajen a lo largo del mismo eje. La célula gaseosa se coloca en el campo de ondas estacionarias que resulta de esta superposición. Entonces se sintoniza el láser a una frecuencia que sea exactamente la mitad de la frecuencia de la transición atómica seleccionada. A la frecuencia apropiada, cada átomo en el gas puede absorber simultáneamente dos fotones que llegan desde direcciones opuestas.

Supongamos que un átomo que se mueve a lo largo del eje óptico en el campo de ondas estacionario encuentra dos fotones que se le acercan en direcciones opuestas. Uno de los fotones tiene un desplazamiento Doppler hacia el azul, esto es, hacia frecuencias más altas, pero el otro fotón está desplazado hacia el rojo en la misma cantidad. Para un átomo estacionario, los dos fotones tienen la misma energía. En todos los casos, la energía o frecuencia total de los dos fotones absorbidos es constante, con independencia de la velocidad del átomo. Cuando la frecuencia del láser se desintoniza ligeramente de la mitad de la frecuencia de la transición, ningún átomo puede absorber los dos fotones que se propagan en direcciones contrarias. (Hay, sin embargo, una pequeña señal de fondo que proviene de átomos en movimiento que absorben dos fotones que se propagan en la misma dirección.) Cuando la longitud de onda del láser coincide con la transición atómica, se observa un rápido aumento en el número de átomos excitados. La señal libre de efecto Doppler se ve aumentada considerablemente, puesto que pueden contribuir a ella todos los átomos de la muestra y no sólo los estacionarios. En este aspecto, el método difiere fundamentalmente de la espectroscopía de saturación y de la espectroscopía de polarización, en las que se selecciona una población de átomos que están libres de efecto Doppler. En la espectroscopía de dos fotones libres de efecto Doppler, se hace que un efecto Doppler compense el otro.

Una transición de dos fotones de particular interés es la transición del estado $1S_{1/2}$ al estado $2S_{1/2}$. Puesto que este último estado tiene una vida media muy larga, se espera que la raya sea extraordinariamente estrecha. Los cálculos teóricos sugieren una anchura natural de la raya de alrededor de un hertz, o menos de una parte en 10^{15}, que haría que la transición fuera una de las resonancias más estrechas conocidas en cualquier sistema físico. El estado $1S_{1/2}$ es en sí mismo interesante, pues su efecto Lamb constituye el mayor de los predichos por la electrodinámica cuántica. Un átomo en el nivel energético $n = 1$ no puede tener momento angular orbital; por tanto, no hay un estado P cercano a partir del que se pueda determinar el tamaño del desplazamiento del nivel por efecto Lamb. El desplazamiento es simplemente un corrimiento del nivel energético $1S_{1/2}$ de la posición que tendría en la ausencia de los efectos predichos por la electrodinámica cuántica. La magnitud del desplazamiento puede determinarse sólo mediante una medición absoluta de alguna componente espectral, tal como la transición $1S_{1/2}$ a $2S_{1/2}$.

Una excitación debida a dos fotones de esta transición fue observada por primera vez en 1974 por Siu Au Lee, Wieman y otros en Stanford. La raya Lyman-alfa tiene una longitud de 1215 angstrom; un fotón con la mitad de esta energía corresponde a una longitud de onda de 2430 angstrom, que aún está en el ultravioleta y más allá del alcance de los láseres sintonizables de colorante. Los investigadores de Stanford emplearon un láser de colorante operando a 4860 angstrom, es decir, dos veces la longitud de onda conveniente. Se hacía incidir la intensa luz de esta fuente sobre un cristal que servía como un duplicador de frecuencia. A causa de la alta intensidad del campo electromagnético aplicado, el cristal no sólo volvía a radiar luz a la frecuencia original sino que también lo hacía en sus armónicos, incluyendo el segundo a 2430 angstrom. Alrededor de un 2 por ciento de la energía aparecía a esta longitud de onda.

La radiación ultravioleta con la frecuencia duplicada era reflejada por un espejo para formar una onda estacionaria dentro de un tubo de descarga gaseosa de baja presión. Los átomos que absorben dos fotones vuelven al estado fundamental emitiendo un único fotón en el ultravioleta lejano con una longitud de onda de 1215 angstrom. Estos

fotones escapaban a través de una ventana lateral de la cámara y eran detectados por un fotomultiplicador.

Las transiciones $1S_{1/2}$ a $2S_{1/2}$ de hidrógeno y de deuterio fueron medidas por este método. Cada una de estas rayas tiene un gran desdoblamiento hiperfino, que quedaba bien resuelto en los espectros de dos fotones. Las transiciones entre los dos estados hiperfinos en el hidrógeno son las responsables de la radiación de 21 centímetros que es una importante señal en radioastronomía. Hay también un gran corrimiento entre las rayas del hidrógeno y las rayas del deuterio, causado principalmente por la diferencia en la masa nuclear. El desplazamiento isotópico fue determinado con una precisión 1000 veces mejor que las mediciones anteriores. Una nueva mejora de un factor 10 permitiría obtener un nuevo valor para la relación de la masa del protón a la del electrón.

El mayor interés en el estado $1S_{1/2}$ es la determinación del efecto Lamb del estado fundamental. En la década de 1950 Herzberg, que estaba entonces en el observatorio de Yerkes de la Universidad de Chicago, pudo detectar este efecto, que midió su valor con una precisión de sólo un 14 por ciento. La espectroscopía de dos fotones daba una oportunidad para una determinación más precisa.

El método más directo de medir el efecto Lamb sería determinar la longitud de onda absoluta de la transición $1S_{1/2}$ a $2S_{1/2}$, pero el valor del desplazamiento Lamb calculado a partir de esta longitud de onda depende también del valor que se adopte para la constante de Rydberg. El grupo de Stanford encontró un método mejor, explotando una coincidencia peculiar pero no accidental: la longitud de onda fundamental del láser de colorante, es decir, 4860 angstrom, casi coincide con la línea Balmer-beta. Si las fórmulas de Bohr fueran correctas, la correspondencia sería exacta: el intervalo de $n = 1$ a $n = 2$ (Lyman-alfa) sería exactamente cuatro veces el intervalo del $n = 2$ al $n = 4$ (Balmer-beta). En realidad, los niveles están divididos y algo desplazados por correcciones debidas a efectos relativistas y a efectos de la electrodinámica cuántica. Sin embargo, las correcciones para los niveles $n = 2$ y $n = 4$ se conocen con gran precisión y por tanto la comparación de las dos transiciones hace posible la determinación del efecto Lamb del estado fundamental.

El grupo de Stanford ha realizado tres mediciones de este tipo, la más reciente y más precisa llevada a cabo por Wieman. La raya de referencia Balmer-beta fue observada por espectroscopía de polarización, mientras el mismo láser medía simultáneamente la línea Lyman-alfa por espectroscopía de dos fotones. El espectro de la raya Balmer-beta no sólo resolvió las componentes de la estructura fina de la raya, sino que también reveló el desdoblamiento de estas componentes en el débil campo eléctrico axial del tubo de descarga gaseosa. En principio, el efecto Lamb del estado fundamental podría determinarse midiendo la diferencia entre la raya Balmer y la raya $1S$ a $2S$ y después comparando los resultados con las predicciones de la teoría de Dirac. Correcciones conocidas se aplicarían a las posiciones de los niveles $n = 2$ y $n = 4$; cualquier diferencia que aún quedara sería atribuida al desplazamiento Lamb del estado $1S$. En la práctica, las cosas no resultaron tan sencillas, pues se debían tener en cuenta los posibles errores sistemáticos y las imprecisiones del aparato. El resultado que se obtuvo fue 8161 ± 29 megahertz. La experimentación aún se halla muy lejos de la exactitud de los cálculos teóricos, que dan un valor de 8149,43 ± 0,08 megahertz, pero representa un paso significativo en comparación con la medición anterior de Herzberg.

Mayor resolución

Hay que vencer muchos y formidables obstáculos para aproximarse a la anchura natural de un hertz de la raya de la transición $1S_{1/2}$ a $2S_{1/2}$. No será suficiente reducir aún más la anchura de banda del láser, pues hay varias otras fuentes que contribuyen al ensanchamiento de la raya. Una de ellas es el efecto Doppler de segundo orden, una consecuencia de la teoría especial de la relatividad. Un átomo en movimiento aparece como si oscilara de forma más lenta que otro que esté en reposo, por lo que el átomo en movimiento tiene efectivamente una frecuencia de resonancia más baja independiente de la dirección en que se mueva. Los haces que se propagan en sentidos opuestos en la espectroscopía de dos fotones no pueden cancelar este efecto, que corresponde a unos 50 kilohertz para átomos de hidrógeno a temperatura ambiente.

Otra fuente de ensanchamiento de la raya es el tiempo de tránsito finito de un átomo moviéndose en el haz del láser. Como un átomo que se mueve queda expuesto a la luz del láser sólo por un breve tiempo, hasta la luz de un láser continuo se percibe por el átomo como un pulso, cuya breve duración limita la resolución potencial del espectro. En principio, ambos efectos pueden reducirse frenando los átomos, o en otras palabras, enfriando el gas. Con relación a esto es interesante que Daniel Kleppner, del MIT, haya probado, recientemente, que el hidrógeno atómico puede enfriarse hasta la temperatura del helio líquido (4,2 grados K) sin condensarse, a pesar de que moléculas de hidrógeno diatómico se condensan a 20 grados K. Dos de nosotros (Schawlow y Hänsch) acaban de revelar que la misma luz del láser puede enfriar un gas a baja temperatura. Esto puede parecer paradójico, ya que una luz intensa usualmente suministra calor, pero debe recordarse que la luz láser, al ser coherente, representa un estado de baja entropía o bajo desorden. Puede lograrse que un estado de baja entropía dé origen a bajas temperaturas.

Varios laboratorios están intentando aplicar la técnica de la espectroscopía de dos fotones a la transición $1S_{1/2}$ a $2S_{1/2}$ del positronio, un "átomo" de tipo hidrogenoide constituido por un electrón y su antipartícula, el positrón. La electrodinámica cuántica predice con gran exactitud las propiedades de este exótico átomo, y no es necesario hacer correcciones debidas a la estructura del núcleo. Sin embargo, la dificultad experimental es grande, pues el positronio tiene una vida media de sólo unos 140 nanosegundos.

El átomo de hidrógeno ha constituido el foco de atención durante casi un siglo, pero la información contenida en este simple sistema de partículas no está agotada. Las teorías de campos unificadores propuestas recientemente, tales como la de Stefen Weinberg, de la Universidad de Harvard, y Abdus Salam, del Imperial College of Science and Technology, predicen efectos sutiles en el hidrógeno y en otros átomos que pueden observarse como pequeños cambios en la polarización de la luz emitida o absorbida. Los cambios son tan pequeños que persiste la duda sobre su posible detección, pero varios grupos de investigadores están intentando medirlos con precisión suficiente para comprobar las teorías. Mientras progrese la tecnología de las fuentes de luz coherente, parece seguro predecir que la exploración del espectro del hidrógeno continuará durante décadas. Hasta ahora, todos los resultados recientes tienden a confirmar las predicciones teóricas. Sin embargo, no está de más recordar que varios de los principales cambios en la física del siglo XX fueron instigados por el descubrimiento de pequeñas discrepancias entre la teoría y la experimentación en este mismo espectro. El átomo de hidrógeno puede aún guardar sorpresas. La sorpresa mayor sería, quizá, que no hubiera ninguna.

Ciencia y sociedad

Genes que violan las leyes de Mendel

En la reproducción sexual, los genes paternos se redistribuyen constantemente, quedando así expuestos de forma equitativa a la severa prueba de la selección natural. Pero hay ciertos genes que "hacen trampa", subvirtiendo el proceso para favorecer su propia supervivencia. El gran papel de la reproducción sexual es de carácter evolutivo: mezcla los genes de los dos progenitores para proporcionar la dotación genética de su descendencia; por tanto, expone al duro examen de la selección natural una máxima variedad de los caracteres y propiedades representados en una especie. Si se pretende que el sistema funcione bien, que la prueba sea justa, la redistribución debe hacerse honestamente. Sin embargo no siempre es éste el caso. Hay genes que hacen trampa, perpetuándose en la población al amañar el proceso de reproducción para que opere en favor suyo.

La trampa tiene lugar en la fase de la reproducción sexual denominada meiosis, la "división reduccional", por la que una célula germinal masculina se divide para dar cuatro espermatozoos, cada uno de los cuales posee la mitad de la dotación normal de cromosomas. Un proceso análogo sucede en la formación del óvulo en la hembra. A la meiosis sigue la fecundación, que restaura el número original de cromosomas, constituyendo el conjunto de estos dos procesos la base física de la herencia mendeliana.

Los cromosomas se presentan en parejas, y por tanto también los genes, que son segmentos del largo filamento de ADN (la parte esencial de los cromosomas). En las células somáticas y germinales de todos los individuos, los miembros de una pareja cualquiera de cromosomas homólogos proviene respectivamente de uno de los dos progenitores. En la meiosis, los miembros de cada par se segregan: uno u otro, al azar, se transmite a un espermatozoo u óvulo; y así, a la descendencia. Los distintos pares de cromosomas se segregan independientemente, de manera que los genes que se encuentran en pares de cromosomas distintos se mezclan por completo en cada meiosis. Es más, existe un proceso habitual, denominado entrecruzamiento, en el que los dos miembros de una pareja se alinean muy próximos, se rompen en puntos correspondientes e intercambian fragmentos equivalentes. En consecuencia, incluso los genes que se encuentran en el mismo cromosoma no están condenados a permanecer juntos, sino que participan de la redistribución meiótica.

La ventaja evolutiva de esta redistribución es que permite reunir los genes de distintos individuos y separar los que han estado juntos. Mezclando así los genes en cada generación sexual, la herencia mendeliana pone a prueba a todos los genes en numerosas combinaciones. Los genes que incrementan la capacidad de supervivencia, de fertilidad y de conservación de la descendencia quedan retenidos por el filtro de la selección natural; los genes menos eficaces se pierden. La evolución implica poner a prueba continuamente combinaciones de genes, y conservar las que aumentan el éxito en la reproducción, y así las especies van mejorando su adaptación.

La meiosis asegura un test escrupulosamente honesto para cada combinación de genes, al dar a cada gen la misma probabilidad de transmitirse a la generación siguiente. Los genes que hacen trampa en la meiosis, debilitan el sistema por medio de la estratagema de reducir su imparcialidad. Estos genes afectan normalmente a otros procesos además de la meiosis; al igual que la mayoría de nuevos genes mutantes, son siempre perjudiciales; por ejemplo, uno de estos genes hallado en poblaciones de ratones causa anormalidades en la cola, además de alteraciones en la meiosis. Se han descubierto otros genes "tramposos" en el maíz, el lirio, el tabaco, en *Trillium*, en el centeno, mosquitos, saltamontes y en la clásica especie favorita de los genéticos: la mosca de la fruta *Drosophila melanogaster*. Un conjunto bien conocido de genes tramposos son los que producen el carácter *"distorsionador de la segregación"* ("segregation distorter"), SD, en las *Drosophila*.

La historia empieza en 1956 cuando el SD fue descubierto por Yuichiro Hiraizumi, quien por entonces estaba investigando ciertos genes del cromosoma número II de poblaciones naturales de *Drosophila* que afectan a la viabilidad. Cruzó machos en los que un miembro de cada par de estos cromosomas era del "tipo salvaje" procedente de la naturaleza, y el otro provenía de una cepa del laboratorio, y fijó la herencia de los cromosomas observando la distribución de determinados genes marcadores que afectan a la pigmentación de los ojos. El ojo normal de *Drosophila* tiene dos pigmentos, uno cinabrio (un color escarlata claro, como el del mineral del que toma el nombre) y otro marrón. Conjuntamente dan al ojo su color normal rojo obscuro. En una forma mutante no se produce pigmento marrón: el ojo es, por tanto, de color cinabrio y la mutación se denomina *cn*. En otra mutación, *bw* (del inglés "brown"), queda suprimido el pigmento de color cinabrio, y el ojo es marrón. Cuando se presentan ambas mutaciones, como era el caso en la cepa de laboratorio de Hiraizumi, el ojo no está pigmentado y se ve por ello blanco. Todos los mutantes son recesivos, es decir, que muestran su efecto sólo si está ausente el gen normal, pues éste lo enmascararía.

Hiraizumi cruzó machos híbridos de ojos rojos (con un cromosoma del laboratorio que contenía las mutaciones y uno natural) con hembras que tenían los dos cromosomas del laboratorio y eran, por tanto, de ojos blancos. Dado que normalmente no hay recombinación en los machos de *Drosophila*, todos los cromosomas transmitidos a la descendencia llevaban, bien los dos genes mutantes de la coloración de los ojos, bien ninguno de ellos, de forma que, de acuerdo con las leyes de la herencia mendeliana, una mitad de la descendencia debería tener ojos blancos y la otra ojos rojos. Hiraizumi observó la distribución 50:50 esperada en la descendencia de unos 200 cruzamientos. Sin embargo, seis de los cruzamientos produjeron resultados muy extraños. No sólo la mitad de los cien y pico descendientes de cada cruzamiento eran de ojos rojos, sino que lo eran casi todos. La proporción de ojos rojos iba del 95 al 100 por ciento y era de un 99 por ciento o más en la mayoría de los casos.

¿Qué había inducido a seis cromosomas, procedentes cada uno de ellos de un macho distinto de la población salvaje, a comportarse de forma tan poco usual? La posibilidad de una muerte selectiva de descendientes de ojos blancos en los primeros estadios del desarrollo embrionario se eliminó rápidamente. Hiraizumi demostró también que lo que estaba desviando la proporción en el color de los ojos actuaba sólo en el curso de la formación de los espermatozoos, no en el de la formación del óvulo.

En casi todas las poblaciones naturales de *Drosophila* estudiadas, procedentes de muchas partes del mundo, del 1 al

5 por ciento de los cromosomas n.º II portaban el carácter SD. Estos cromosomas tienen otra peculiaridad que debe ser mencionada: prácticamente en todos los casos tienen una inversión, y normalmente más de una. Las inversiones (regiones en que los genes se hallan dispuestos en un orden inverso al normal) no son infrecuentes en *Drosophila*, pero es poco usual hallar dos o más en un mismo cromosoma. Debía haber alguna explicación para esta asociación entre el carácter SD y las inversiones múltiples.

Algunos años antes del experimento de Hiraizumi, Laurence M. Sandler y Edward Novitski sugirieron, en base a ciertas indicaciones experimentales, que las leyes de la meiosis podían ser a veces violadas. Denominaron "deriva meiótica" ("meiotic drive") a este proceso y consideraron algunas de sus posibilidades teóricas.

El cromosoma SD no repercute de forma visible en las moscas portadoras, el único efecto apreciable es una distorsión en la proporción de los tipos de la descendencia. En los complicados experimentos requeridos para analizar el sistema, los cromosomas van siempre marcados con genes mutantes visibles, como los genes del color de los ojos, de forma que el cromosoma puede ser seguido a lo largo de la compleja serie de cruzamientos.

Un hecho que dificulta la comprensión de cómo el cromosoma SD da lugar a este cambio tan fundamental en el comportamiento meiótico es que el cromosoma debe inactivar de alguna forma precisamente a aquellos espermatozoos que no lo contienen. ¿Cómo ocurre?

Sandler e Hiraizumi sugirieron muy pronto una posibilidad: mientras los cromosomas homólogos se mantienen apareados durante la meiosis, el cromosoma SD debe hacer algo a su compañero (y rival), que más tarde causa una disfunción en el espermatozoo que recibe el cromosoma normal. En principio sugirieron que el SD podría romper al otro cromosoma. Sin embargo, W. J. Peacock fue incapaz de confirmar la hipótesis de rotura cromosómica. Observaciones microscópicas mostraron que los cromosomas de heterozigotos SD (células con un cromosoma SD y otro normal) atravesaban la meiosis sin daño alguno, de forma que el efecto perjudicial del SD debería ser más sutil que el de una rotura. Peacock y Erickson propusieron una alternativa más ingeniosa. Algún tiempo antes, Novitski e Iris Sandler habian sugerido que, incluso en machos normales, sólo dos de los cuatro espermatozoos producidos en una meiosis eran funcionales. Peacock y Erickson pensaron que los cromosomas SD deberían aprovecharse de ello, consiguiendo incluirse de alguna manera en los espermatozoos que estaban destinados a ser funcionales.

Podría parecer sencillo distinguir entre las dos hipótesis: disfunción inducida del espermatozoo, e inclusión preferente. Si el SD causara la disfunción de la mitad de los espermatozoos, entonces los machos con un cromosoma SD deberían producir sólo la mitad de espermatozoos funcionales que los machos normales, y la producción reducida de espermatozoos se reflejaría en una reducción de la fertilidad. El problema es que el número de espermatozoos producidos normalmente por una mosca macho es mucho mayor que el requerido para fecundar todos los óvulos de una hembra; en consecuencia, el deterioro de incluso la mitad de los espermatozoos podría no reducir la fertilidad.

Los investigadores Daniel L. Hartl, Hiraizumi y James F. Crow hallaron un camino para evitar este problema. Cruzando machos muy jóvenes (que producen menos espermatozoos que los machos maduros) o cruzando un macho con varias hembras durante algunos días hasta agotar la provisión de espermatozoos del macho, hicieron de la producción de espermatozoos el factor limitante que determinara el número de descendientes. Independientemente, Benedetto Nicoletti y Gianni Trippa llevaron a cabo experimentos similares. Ambos grupos llegaron a las mismas conclusiones. Los machos que portaban un cromosoma SD producían, desde luego, menos descendientes que los machos normales, y no el mismo número, como hubiera sido el caso si el mecanismo fuera el de una inclusión preferencial. Además, el número de descendientes quedaba reducido justo en la proporción que era de esperar en el caso de que el cromosoma SD causara una disfunción de los espermatozoos que recibían el cromosoma normal.

La prueba de una disfunción inducida por el SD vino de estudios de la maduración de los espermatozoos efectuados mediante microscopio electrónico. Nicoletti refirió que alrededor de la mitad de los espermatozoos de machos SD tenían la cola con una apariencia anormal. Más tarde, Peacock, Kiyoteru T. Tokuyasu y Robert W. Hardy, de la Universidad de California en San Diego, confirmaron los hallazgos de Nicoletti y revelaron detalles adicionales. Demostraron que, aunque normalmente todos los núcleos de un haz de 64 espermatozoos inmaduros (derivados de un grupo de 16 células germinales por dos divisiones meióticas) se hacen menores y densos para formar la cabeza del espermatozoo, en las moscas que portaban un cromosoma SD y uno normal sólo se condensaban de forma correcta la mitad de los núcleos. Los espermatozoos que no lograban desarrollarse correctamente eran con claridad los que no contenían el cromosoma SD.

La deducción obvia que surge de estos resultados es que el cromosoma SD daña de alguna manera a su homólogo no-SD y lo hace incapaz de llevar a cabo su función usual. Sin embargo, esto no puede ser así, ya que la incapacidad de desarrollarse no puede ser sencillamente la consecuencia de una incapacidad de llevar a cabo un proceso normal requerido para la maduración, ya que sabe, desde hace tiempo, que el funcionamiento de un espermatozoo no depende, en absoluto, de su contenido cromosómico.

El SD provoca algo en su homólogo que hace que éste a su vez produzca la disfunción del espermatozoo. El SD se perpetúa a sí mismo al inducir a su compañero a destruirse. Se desconoce cómo el homólogo normal es instruido por el SD a conducirse incorrectamente, así como la naturaleza de este comportamiento maligno.

El cromosoma SD debe inducir a su homólogo a volverse activamente perjudicial a la función del espermatozoo.

Cualquiera que sea la razón evolutiva que ha determinado que no funcionen los genes en los espermatozoos, este hecho dificulta el surgimiento de sistemas dañinos análogos al SD, y a la vez impide una competición, negativa o irrelevante, entre espermatozoos portadores de genes distintos.

No hay forma de reconocer una mosca SD por su apariencia; el SD afecta a la proporción de la descendencia, y las proporciones distorsionadas pueden apreciarse sólo si el cromosoma está marcado con genes mutantes. El mapaje genético por métodos tradicionales de recombinación se ve complicado por las secuencias génicas invertidas, que tienden a suprimir el entrecruzamiento. Los experimentos han demostrado que el principal efecto de segregación-distorsión está causado por dos genes que se hallan muy próximos, pero separados por el centrómero (aquel punto al que están unidas las fibras que separan los cromosomas hermanos durante la división celular).

Dado que el cromosoma SD es transmitido a casi toda la descendencia, en vez de sólo a la mitad, podría extenderse como un reguero de pólvora por la población, y reemplazar en seguida a su

homólogo. Se conoce ya bastante bien por qué el cromosoma SD no alcanza de hecho una frecuencia del 100 por ciento: un macho homozigoto para SD (con dos cromosomas SD) es casi estéril. En algunas cepas, estas moscas ni siquiera logran sobrevivir. En una población de moscas, la tendencia del heterozigoto SD a aumentar su frecuencia mediante distorsión de la segregación queda contrarrestada por la tendencia del homozigoto SD a la esterilidad o muerte.

Un freno al incremento del SD lo proporcionan varios "genes modificadores" que rebajan el grado de distorsión; estos genes se hallan dispersos por los cuatro cromosomas de *Drosophila* y se encuentran en casi todas las poblaciones naturales de moscas.

Podemos reconstruir lo que pasó. De alguna forma, hace mucho tiempo, apareció el cromosoma *S-R* (*S*, distorsionador de la segregación y *R*, respondedor). Primero aumentó su frecuencia rápidamente. El aumento fue detenido, no obstante, por un incremento de cromosomas normales resistentes, es decir, de cromosomas con la dotación S^+-*R*. Se originó así una competición en tres direcciones: la tendencia a incrementar del cromosoma *S-R* era contenida por el S^+-*R*, que anulaba la ventaja meiótica del primero, y por supuesto el cromosoma S^+-*R* aventajaba al normal, S^+-R^+, que difícilmente logra ser transmitido si se encuentra frente al *S-R*.

El cromosoma SD parece estar controlado por la presencia del cromosoma S^+-*R* y de los genes modificadores, de forma que la frecuencia del SD en una población no llegue a ser lo suficientemente elevada como para causar daño. El sistema SD, y otros sistemas de deriva meiótica que se pueden encontrar en la naturaleza, son aquellos a los que se han ido ajustando las poblaciones. Las poblaciones en las que no se han desarrollado un ajuste análogo deben sencillamente haberse extinguido.

¿Qué pasaría si un sistema como el SD apareciera en uno de los cromosomas que determinan el sexo? Por ejemplo, un macho normal produce en principio un número equivalente de espermatozoos que portan el cromosoma *X* y el cromosoma *Y*, de forma que cuando el espermatozoo fecunda el óvulo portador de un *X*, se producen hembras *XX* y machos *XY* en igual cantidad. Ahora bien, si se localizara un complejo SD en el cromosoma *Y*, cabría esperar que casi el 100 por ciento de la descendencia fuesen machos. No se conoce tal cromosoma *Y* en *Drosophila*, pero Terrence W. Lyttle, diseñó el camino para obtener uno. Por medio de radiación provocó un traslado de los segmentos cromosómicos que sujetaban el fragmento SD al cromosoma *Y*. La distorsión originada por el SD hace que los machos que poseen esta traslocación produzcan un exceso de espermatozoos con cromosoma *Y*, con lo que casi toda la descendencia son machos. Una población afectada por un sistema así se extinguirá en un par de generaciones por falta de hembras.

Sin duda, un sistema de deriva meiótica, que ya es malo hallándose en un cromosoma no sexual, puede significar un desastre si está en el cromosoma *Y*. W. D. Hamilton, de la Universidad estatal de Michigan, ha sugerido que el hecho de que el cromosoma *Y* esté en muchas especies bastante desprovisto de genes puede deberse a que la falta de genes previene la aparición de mutantes creadores de deriva meiótica, que pueden aniquilar la población.

La herencia mendeliana es un dispositivo maravilloso para hacer de la evolución por selección natural un proceso eficiente. Podría tenerse por el mejor sistema diseñable (teniendo en cuenta las limitaciones mecánicas impuestas por el hecho de estar los genes unidos en cromosomas) para someter a cada gen a una prueba minuciosa en combinación con muchos otros genes. El sistema mendeliano trabaja con máxima eficacia sólo si es escrupulosamente justo para todos los genes. Sin embargo, se halla en constante peligro de ser perturbado por genes que subvierten el proceso meiótico en beneficio propio. Si estos genes tienen efectos dañinos (como el de la esterilidad o mortalidad inducida por el SD homozigoto), la población resultará inmediatamente debilitada. Aun cuando los genes "tramposos" no sean en sí perjudiciales, inhiben el proceso evolutivo al reducir su eficacia. Existen muchas sutilezas en la meiosis y la espermatogénesis cuya finalidad es aparentemente impedir esta trampa. Y ya algunos genes han logrado vencer al sistema.

La absolución de Ptolomeo

En diciembre de 1977 esta misma sección aportó una serie de argumentos destinados a mostrar que Claudio Ptolomeo, el astrónomo griego que floreció en el siglo segundo en Alejandría, fue un impostor. Abrió el proceso Robert R. Newton, del Laboratorio de Física Aplicada de la Universidad Johns Hopkins, en un libro titulado *The Crime of Claudius Ptolemy (El crimen de Claudio Ptolomeo)* editado por la Johns Hopkins University Press.

La acusación de Newton no es ninguna banalidad. Ptolomeo está considerado como el mayor astrónomo de la antigüedad. Su *'E Mathematike Syntaxis (Tratado de matemática)*, al que los intelectuales árabes del siglo noveno dieron el nombre de *Almagest (El más grande)*, presenta una teoría global de los movimientos planetarios que fue aceptada universalmente durante 1400 años. En base de esta teoría, Ptolomeo desarrolló un elaborado sistema matemático para predecir las posiciones futuras de los planetas. En la *Syntaxis*, Ptolomeo también reunió el catálogo estelar más completo de la antigüedad. La *Syntaxis* contiene numerosas referencias a los astrónomos que le precedieron, en particular a Hiparco (que editó el primer catálogo estelar); en muchos casos, el testimonio de Ptolomeo constituye la única fuente de información de que disponen los historiadores sobre los albores de la astronomía griega y sobre la cronología babilónica. Por ejemplo, incluyen la única lista con la duración de los reinados de los reyes babilónicos que se posee hoy en día.

Newton acusa al astrónomo de la fabricación sistemática de datos que soportasen su teoría de los movimientos planetarios. "Esto hace de Ptolomeo, escribe Newton, el impostor con más éxito de la historia de la ciencia", y el autor de un libro que "ha hecho más daño al saber que cualquier otro trabajo nunca escrito". El fundamento de los argumentos de Newton es el análisis estadístico; demuestra que la precisión de ciertas observaciones astronómicas que Ptolomeo afirmó haber hecho es tan elevada que, en algunos casos, las probabilidades de que no pudo hacerlas con los instrumentos que describe son de mil millones a una. Por contra, allí donde las observaciones de Ptolomeo son inexactas de acuerdo con la moderna teoría, Newton sostiene que debería haber sido capaz de realizarlas con una mayor exactitud empleando los instrumentos que cuenta que empleó. También aquí ha calculado Newton las probabilidades en contra de que Ptolomeo hubiese cometido tales errores. Al obtener una relación entre dichas probabilidades, del orden de 10^{92}: 1, Newton concluye que las "observaciones ptolomaicas" son puro fraude.

En un artículo que aparecerá en *The American Scholar*, Noel M. Swerdlow, de la Universidad de Chicago, mantiene que el análisis estadístico de Newton no es válido. Para calcular la probabilidad de un determinado suceso que ocurre un cierto número de veces, Newton se apoya a menudo en lo que se conoce con el nombre de regla del producto: la

probabilidad de un único acontecimiento se multiplica por el número de pruebas. Por ejemplo, la probabilidad de obtener un uno al lanzar un dado es un sexto, es decir 1/6. La probabilidad de obtener dos unos sucesivos es 1/6 veces 1/6, o sea 1/36, y la probabilidad de sacar tres veces seguidas uno es 1/6 veces 1/6 veces 1/6, o lo que es lo mismo, 1/216.

Para aplicar la regla del producto es necesario conocer la probabilidad de un suceso y el número total de intentos. Pero, además, esta regla sólo se puede aplicar cuando los sucesos son independientes entre sí. En otras palabras, para que la regla del producto sea aplicable, el que se dé o no un cierto acontecimiento en un intento no puede afectar la probabilidad de que este mismo suceso se obtenga o no en las demás pruebas. Esta condición se cumple para el caso del dado: si en una tirada aparece el uno esto no afecta la probabilidad de que en otras tiradas aparezca un uno.

Según Swerdlow, la regla del producto no se verifica para el caso de las antiguas observaciones astronómicas, tales como la de Ptolomeo, debido a que no se dan ninguna de las condiciones requeridas para su aplicación. Sencillamente, Newton no tenía manera de determinar la probabilidad de que alguna de las observaciones de Ptolomeo tomase un determinado valor, y no podía conocer de ninguna forma el número total de observaciones de las cuales Ptolomeo seleccionó las que nos dejó en sus escritos. Además, Newton no podía saber si estas observaciones son o no independientes entre sí. Por tanto, concluye Swerdlow, Newton aplica erróneamente los métodos estadísticos cuando empieza suponiendo una probabilidad inicial de uno en 10, o 1/10, de que una observación de un eclipse lunar no sea fraudulenta y la multiplica hasta llegar a $1/10^{12}$ para calcular la probabilidad de que, de doce observaciones de un eclipse lunar, ninguna sea fraudulenta.

Las observaciones de Ptolomeo confirman sus cálculos teóricos de una forma tan exacta que desde el punto de vista de la ciencia actual sugieren que hayan sido fabricados. Victor E. Thoren, de la Universidad de Indiana, y Owen J. Gingerich, de la Universidad de Harvard, han señalado de forma independiente que este tipo de precisión es del todo comprensible desde el punto de vista de la ciencia antigua. Las personas dedicadas a la astronomía en tiempos de Ptolomeo eran matemáticos, que se interesaban más por la demostración, el rigor, la lógica y la coherencia interna que por la precisión de las observaciones. La ética establecida en la ciencia antigua era dar cuenta sólo de aquellas observaciones que mejor confirmasen la teoría, y despreciar las demás. Esta costumbre explica la gran concordancia entre las observaciones de Ptolomeo y su trabajo teórico. No fue hasta el desarrollo de los métodos estadísticos y probabilísticos a finales del siglo XVIII cuando los filósofos naturalistas empezaron a recopilar todas las observaciones aleatorias, debido a que sólo entonces tenían técnicas para interpretar grandes volúmenes de datos imprecisos al calcular medias, medianas, desviaciones típicas y otros parámetros similares. En resumen, según Swerdlow, Thoren y Gingerich, la acusación de Newton contra Ptolomeo cae por su base debido a que se sostiene en un defectuoso análisis estadístico y no tiene en cuenta los métodos empleados por la astronomía antigua.

Correa sinfín

Para describir un río no hay necesidad de especificar el estado de cada molécula de agua. Pueden ignorarse los movimientos de las distintas moléculas porque no influyen en las ondas y corrientes hidrodinámicas. Sin embargo, bajo ciertas condiciones, los sucesos microscópicos pueden tener consecuencias macroscópicas. Para el agua estas condiciones son una temperatura cercana a los 647 grados Kelvin y una presión próxima a las 217 atmósferas. Esa combinación de temperatura y presión define el punto crítico del agua, donde el líquido y el vapor tienen la misma densidad y desaparece la distinción entre las dos fases. Cerca del punto crítico, las fluctuaciones aleatorias de densidad, que siempre están presentes, pero que por lo general están confinadas en agregados de unas pocas moléculas, crecen enormemente. El agua se vuelve lechosa debido a que las fluctuaciones dispersan la luz. En el punto crítico, la escala de fluctuaciones se hace infinita, y la perturbación de una única molécula puede cambiar la fase de todo el volumen. Cualquier descripción del agua bajo estas condiciones debería tener en cuenta su estructura a todos los niveles posibles, desde la escala de moléculas individuales hasta la totalidad del fluido. Se ha demostrado que formular una descripción así constituye un problema irresoluble.

Un número sustancial de problemas relevantes en física (y por lo menos algunos más en otros campos) se han resistido a ser solucionados por la misma razón. Así, la cuestión del comienzo de la magnetización espontánea cuando se enfría un material ferromagnético por debajo del punto crítico conocido por el nombre de temperatura de Curie. Muy por encima de este punto, los spin de los electrones desemparejados, que son los imanes fundamentales del sistema, están orientados al azar. Al acercarse a la temperatura de Curie, los spin comienzan a alinearse, primero en pequeña escala e involucrando sólo a los átomos vecinos y después a escalas cada vez mayores. A la temperatura de Curie, el cambio de orientación de un solo spin podría alterar la dirección de la magnetización que a la larga adquiera todo el material.

Las transiciones a estados superconductores y a estados superfluidos, las transiciones entre fases ordenadas y desordenadas en las aleaciones, el desarrollo de la turbulencia en los fluidos, la conformación de polímeros de cadena larga, la difusión de un líquido a través de una matriz sólida y las interacciones de partículas elementales en una teoría cuántica de campos son ejemplos de problemas relacionados con los anteriores. La raíz de la dificultad con que nos hallamos a la hora de explicar teóricamente esos fenómenos estriba en la necesidad de trabajar a la vez con muchas escalas de longitud y de energía. En los últimos años ha empezado a abrirse paso un nuevo enfoque de estos problemas: el llamado grupo de renormalización. La técnica, en síntesis, trata cada escala de longitud sucesivamente, comenzando por la menor.

El procedimiento matemático de renormalización se ideó en la década de 1940 como un peldaño en la formulación de la electrodinámica cuántica, la teoría que describe las interacciones entre una partícula cargada eléctricamente y un campo magnético. En este caso la dificultad era una infinidad de escalas de energía surgidas al calcular la masa o la carga de un electrón. Durante la década de 1950 se puso de manifiesto que la esencia del procedimiento de renormalización podía generalizarse a una gran familia de transformaciones matemáticas; éstas constituyen el grupo de renormalización. Los métodos de resolución más recientes de problemas basados en el grupo de renormalización difieren significativamente de los anteriores en el procedimiento matemático y en el punto de vista físico. En realidad, los dos métodos son lo bastante distintos como para haber sido denominados jocosamente "antiguo" y "moderno". El método moderno fue introducido en 1971 por Kenneth G. Wilson, de la Universidad de Cornell, que basó parcialmente su trabajo en ideas anteriores de Leo P. Kadanoff, de la Universidad de Chicago.

Para explicar cómo actúa el grupo

de renormalización, consideremos de nuevo un material ferromagnético próximo a la temperatura de Curie. Examinado a escala microscópica, el imán es una red de átomos tal que cada uno de ellos tiene asociado un vector de spin que puede apuntar hacia cualquier dirección. Se supone que la red es de tamaño infinito, lo que a escala microscópica no es ninguna hipótesis irracional. Como primer paso del análisis del sistema, se divide la red en bloques de unos pocos átomos cada uno. A continuación, se sustituyen todos los spin de un mismo bloque por un spin promedio en todo él; de esta forma queda reducido el número de spin por unidad de volumen. La etapa final consiste en volver a aumentar la escala de cada bloque a las dimensiones de la red original y ajustar la magnitud del spin. Mediante esta serie de operaciones queda definida una nueva red de átomos que refleja la distribución de los spin en una escala ligeramente mayor.

Para obtener resultados significativos hay que repetir esas operaciones muchas veces. (Los métodos relacionados con el grupo de renormalización exigen complejas habilidades en los computadores.) Durante las primeras iteraciones, la configuración de los spin varía espectacularmente de una generación a la siguiente. Sin embargo, al seguir la computación las variaciones se amortiguan y, a la larga, la configuración se hace estable; se dice entonces que el sistema ha alcanzado un punto fijo. Se repite luego todo el proceso con unas condiciones iniciales ligeramente distintas. Habida cuenta de la naturaleza del punto fijo y de la forma en que el sistema se acerca a él, pueden determinarse ciertas propiedades del imán tales como su temperatura crítica.

El programa para describir una red a escalas progresivamente mayores se define de un modo matemático, pero puede dársele una justificación intuitiva. Esta estrategia conduce a una descripción correcta porque, según ha señalado Wilson, relaciona sucesos en escalas de longitud adyacentes, muy juntas en la naturaleza. Una fluctuación momentánea en la escala atómica de uno o dos angstrom (un angstrom equivale a 10^{-8} centímetros) puede condicionar intensamente las fluctuaciones con una longitud característica de dos a cuatro angstrom, y éstas afectar a las fluctuaciones a una escala de entre cuatro y ocho angstrom, y así sucesivamente.

La comunicación a través de una cascada de escalas de longitud da cuenta de una de las propiedades más sorprendentes de los fenómenos críticos: la extrema sensibilidad a las pequeñas perturbaciones. En un material ferromagnético cerca del punto de Curie, un pequeño cambio de temperatura sólo puede alterar directamente la dirección de unos pocos spin; ahora bien, cuando estos ligeros cambios se propagan a través de la cascada de escalas de longitud pueden desencadenar efectos macroscópicos.

Aunque así se amplifican algunas características, debemos destacar también que otras se ven fuertemente atenuadas. Por ejemplo, dos materiales ferromagnéticos pueden tener estructuras atómicas diferentes, y, sin embargo, al aumentar la escala de longitudes pueden desaparecer de forma gradual las diferencias estructurales. A la temperatura de Curie, el comportamiento macroscópico de los dos cuerpos ferromagnéticos es idéntico. Dentro de un contexto mayor, la moderna teoría del grupo de renormalización parece confirmar una hipótesis algo anterior conocida con el nombre de universalidad del punto crítico, que afirma que cualquier sistema cerca de su punto crítico, sin importar, su composición, debe caer en una de las pocas categorías de comportamiento crítico posibles. La categoría aplicable queda determinada por el número de dimensiones necesario para describir el sistema y por sus propiedades de simetría. Sistemas que, en apariencia, son completamente distintos pueden pertenecer a la misma categoría y, por ende, deben ser formalmente idénticos. Por ejemplo, una gráfica de la densidad del agua cerca del punto crítico es esencialmente igual a otra gráfica de la magnetización del níquel cerca de la temperatura de Curie.

Aunque está ampliamente reconocida la posible potencia del grupo de renormalización, sus rendimientos, por ahora, son modestos. El comportamiento crítico de los fluidos y de las sustancias ferromagnéticas sólo se ha resuelto para sistemas modelo, no para sustancias reales. A pesar de que los sistemas modelo son versiones del problema real altamente simplificadas, en algunos casos ni tan siquiera éstos han sido resueltos por cualquier otro método. Los estudios experimentales recientes de fluidos cerca del punto crítico tienden a confirmar los resultados del análisis mediante el grupo de renormalización.

El mayor éxito del grupo de renormalización es una solución precisa de un problema de física del estado sólido llamado efecto Kondo. Este efecto se observa al contaminar un metal no magnético, como el cobre, con trazas de un metal magnético, como el hierro. A baja temperatura, aparecen anomalías en las propiedades eléctricas del metal impurificado. El afán por explicar estas anomalías dio pie a un considerable trabajo durante la década de 1960, no porque el problema fuese de suyo importante, sino porque era como una espina incitadora: la solución parecía estar al alcance de la mano. En realidad el efecto Kondo conlleva un problema de muchas escalas de energía. Es necesario evaluar las interacciones de las impurezas magnéticas con los electrones de conducción del metal, que pueden tomar cualquier energía dentro de una banda ancha y esencialmente continua. Por fin, Wilson presentó una solución basada en el grupo de renormalización, en 1974.

Un problema de indiscutible interés, al que Wilson y otros han dedicado un notable esfuerzo, es la descripción de las interacciones entre quarks, que son los supuestos constituyentes de los protones, neutrones y otras muchas partículas fundamentales. Se ha esbozado una teoría provisional sobre estas interacciones, que se funda en la electrodinámica cuántica y recibe el nombre de cromodinámica cuántica. Sin embargo, esta teoría sólo es una herramienta predictiva, útil bajo un restringido conjunto de condiciones; pero hay muchas e importantes propiedades de los quarks y de las partículas que todavía no pueden describirse con ella. De nuevo, la dificultad radica en tener que tratar muchas escalas de energía o de longitud. Cuando se examinan quarks a altas energías y distancias cortas (por ejemplo, en el interior del protón), parece que sólo estén débilmente unidos entre sí. Sin embargo, a bajas energías y largo alcance, se unen con tal fuerza que no parece exista modo alguno de separarlos. La cromodinámica cuántica explica satisfactoriamente el comportamiento de los quarks a corto alcance, pero todavía no se ha encontrado un método para calcular las interacciones de largo alcance.

Obviamente, el grupo de renormalización es un método candidato para extender la cromodinámica cuántica a todas las escalas de longitud y de energía. No se han obtenido todavía resultados numéricos, pero se ha sugerido una hipótesis. En la electrodinámica cuántica, la separación de las cargas eléctricas (el proceso llamado ionización) puede contemplarse como una transición de fase, regida por un punto fijo en un espacio abstracto. Para los quarks puede no existir el correspondiente punto fijo o estar en una región accesible sólo con energía infinita. En ambos casos, los quarks estarían confinados en una fase condensada, como el agua que no puede hervir nunca sin importar cuánto se calienta.

La percepción de los blancos y negros de las superficies

El tono gris con que aparece una superficie está relacionado con la distribución de las luces y sombras que se perciben en ella; dicha distribución depende a su vez de la relación existente entre la superficie y el espacio circundante

Alan L. Gilchrist

El cristalino del ojo humano proyecta en la retina una imagen bidimensional del mundo físico, que es tridimensional. En parte porque la imagen retiniana es bidimensional, la mayoría de los investigadores en la percepción del color han supuesto que la percepción del relieve no está relacionada con el mecanismo a través del cual el sistema visual humano determina el color de los objetos. Los experimentos que voy a describir a continuación invalidan esta suposición, pues he descubierto que una modificación de la orientación espacial de una superficie puede hacer variar el color percibido de blanco a negro y de negro a blanco.

Las explicaciones tradicionales que daban cuenta de la percepción del color no concedían ninguna misión especial en este proceso al sistema nervioso central. Dado que este sistema es el responsable de la percepción del relieve, mis estudios le confieren un papel fundamental. Por otra parte, las investigaciones tradicionales han venido ignorando generalmente la percepción de la intensidad y del color de la luz que ilumina las superficies cuyos colores son sometidos a examen. En mis estudios he tratado de corregir estos desequilibrios y en base a la simplificación he trabajado sobre todo con los colores que se ha dado en llamar neutros: el blanco, el gris y el negro.

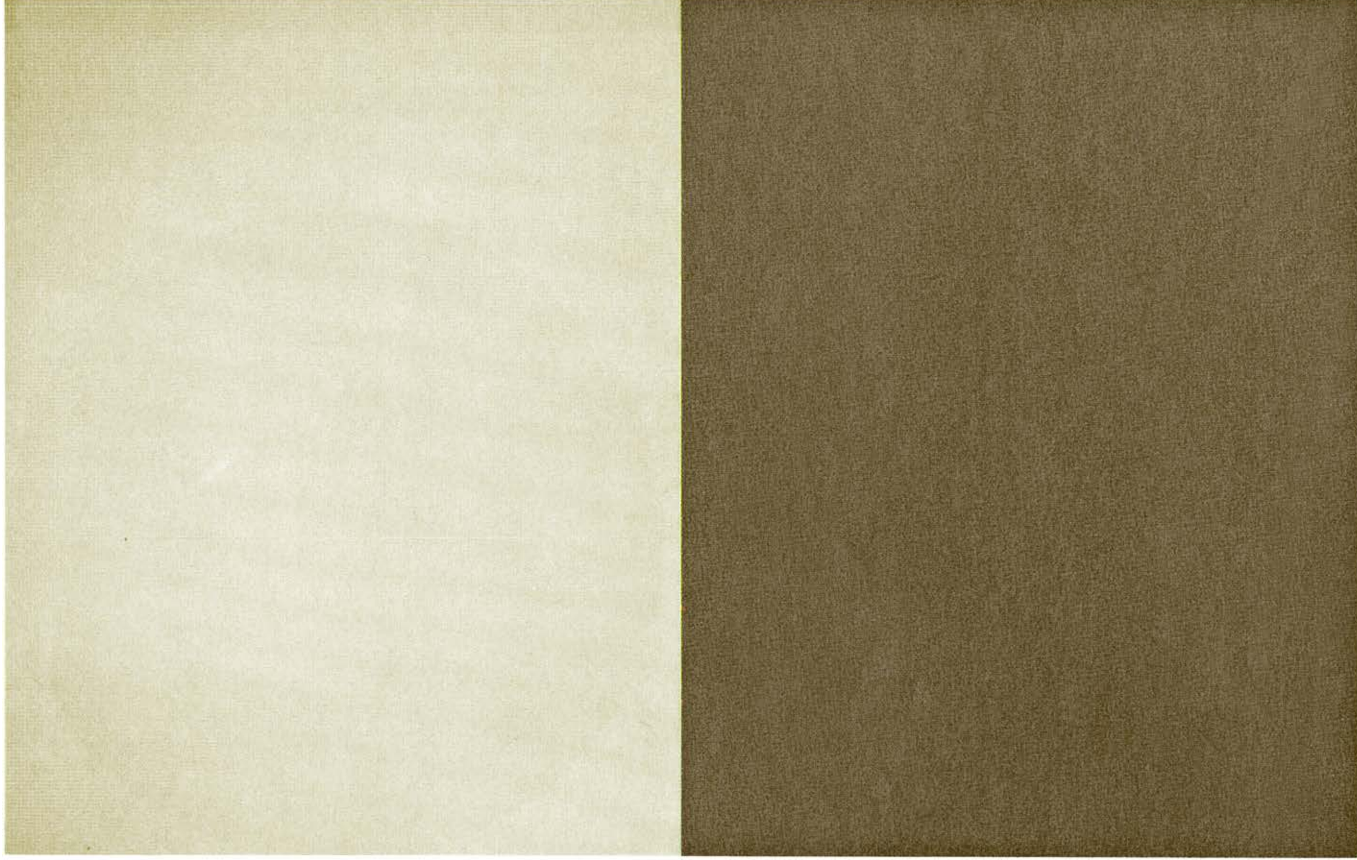
Cuando se hace incidir luz blanca (que es una mezcla equilibrada de todos los colores del espectro) sobre una superficie de un determinado color, la superficie refleja una longitud de onda (un color) más que otros. Esta longitud de onda dominante corresponde al color físico de la superficie. Generalmente la luz que ilumina los objetos en la naturaleza (la luz natural) no es blanca, sino que es una mezcla desequilibrada de varios colores. Una superficie coloreada refleja algunas longitudes de onda más que otras, pero en este caso las longitudes de onda dominantes no corresponden al color físico de la superficie. La mezcla de longitudes de onda en la luz que una superficie refleja al ojo depende no sólo del color de esa superficie, sino también de la mezcla de longitudes de onda en la luz que ilumina la superficie. Por tanto, el sistema visual humano no debería considerarse como un instrumento que mide la longitud de onda y la intensidad de la luz reflejada: tales mediciones revelarían poco acerca del color de una superficie. El sistema visual tiene la extraordinaria capacidad de percibir correctamente el color físico de una superficie a pesar de las amplias variaciones que se puedan producir en la mezcla de longitudes de onda de la iluminación. Este es el fenómeno de la constancia del color.

Las superficies neutras dejan invariante la proporción de la mezcla de longitudes de onda que las iluminan, si bien modifican la intensidad de la luz. Los tonos de gris, desde el blanco hasta el negro, son todos neutros; por consiguiente, reflejan grados distintos de iluminación. Las superficies que reflejan alrededor del 80 o 90 por ciento de iluminación se denominan blancas, mientras que las que reflejan alrededor del 3 al 5 por ciento son llamadas negras. En pocas palabras, la luminosidad o el grado de gris percibido en una superficie neutra se corresponde con su reflectancia: el porcentaje de iluminación que refleja. De nuevo, el sistema visual tiene la virtud especial de determinar la luminosidad de una superficie neutra a pesar de que se puedan producir amplias variaciones en la intensidad de la iluminación. Se trata del fenómeno de la constancia en la luminosidad, que es análogo al de la constancia en el color para superficies no neutras.

Un físico determinaría la reflectancia de una superficie comparando la intensidad de la luz reflejada con la de la incidente. El sistema visual humano es asimismo capaz de determinar reflectancias, pero lo hace comparando la intensidad reflejada por una determinada superficie con la reflejada por las superficies más próximas. Para que el sistema visual humano pueda efectuar tal comparación, tiene que existir un desplazamiento continuo entre la retina y la imagen retiniana. Esto se consigue porque el ojo se está moviendo continuamente hacia adelante y hacia atrás, fluctuando con una frecuencia de entre 30 y 150 ciclos por segundo. Se ha comprobado que el campo visual desaparece en un intervalo de tiempo de uno a tres segundos si el desplazamiento relativo antes mencionado se elimina estabilizando artificialmente la imagen (usando un dispositivo especial que hace que la imagen se forme siempre sobre la retina, aunque el ojo se esté moviendo de atrás adelante). Las fluctuaciones del ojo ponen de manifiesto que las células receptoras de la retina sólo actúan en presencia de una estimulación cambiante.

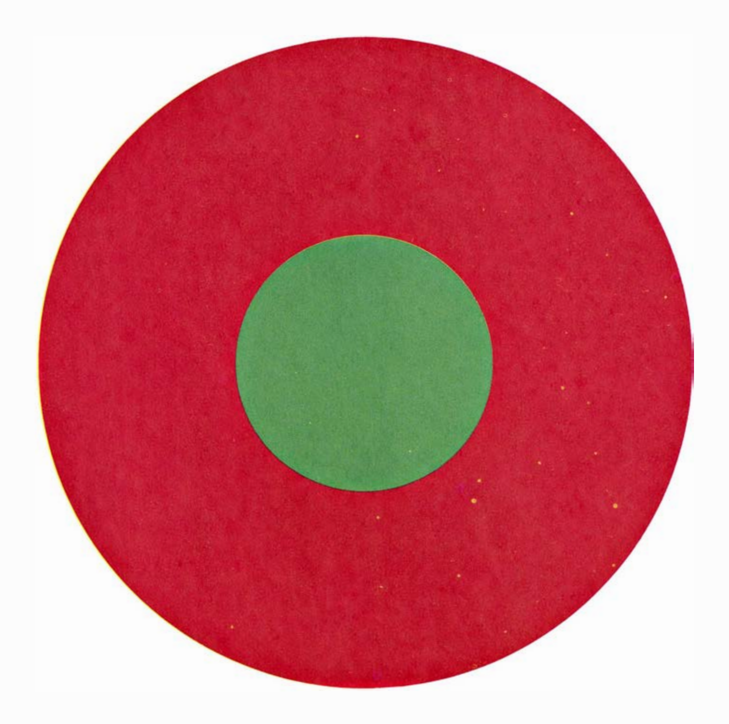
Veamos qué le sucede a una célula receptora individual durante estas fluctuaciones (suponiendo que no existen grandes movimientos voluntarios del ojo). Una célula de la retina que se encuentre situada dentro de una zona de estimulación homogénea, no recibe ningún cambio de estimulación, mientras que una célula situada justo en el límite de esta zona, sí lo recibe.

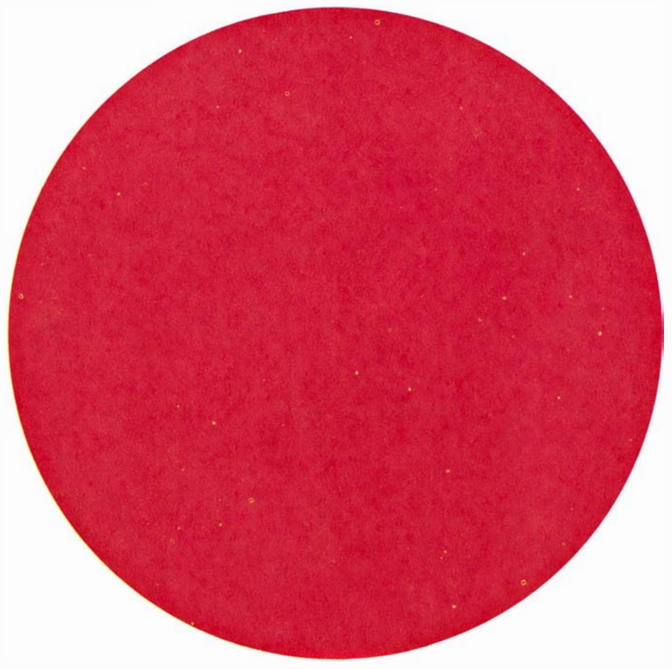
La importancia de todo esto en el proceso de la visión del color fue resaltada de un modo coherente por John Krauskopf, de los Bell Laboratories. Construyó un disco de un determinado color, por ejemplo verde, rodeado por un anillo de otro color, rojo por ejemplo. Dispuso las cosas de tal manera que la fron-



LA ORIENTACION ESPACIAL PERCIBIDA influye en la percepción del color. En la fotografía superior se ve que la parte de la derecha del hueco de la escalera es una superficie blanca en sombra y que la parte de la izquierda es una superficie blanca iluminada. El sistema visual percibe correctamente que el hueco de la escalera no está iluminado uniformemente porque puede percibir que las escaleras hacen una esquina. La fotografía inferior es una vista cercana de la esquina del hueco de la escalera. Se diría que aquí ambas caras forman un plano. Las caras se perciben con la misma iluminación y el sistema visual atribuye la diferencia en el tono de gris a las caras en sí. La cara que está en la sombra no se ve de color blanco.

LIMITES, o diferencias, de iluminación. Son de dos tipos, intrínsecos y proyectados. Los límites intrínsecos son el resultado de los cambios en la orientación espacial de las superficies; los límites proyectados son el resultado de las sombras. En esta fotografía, ambos tipos de límites están presentes con abundancia. Por ejemplo, hay límites intrínsecos allí donde las paredes hacen esquinas para formar los soportes de las ventanas. Las fronteras de las zonas de luz trapezoidales que se forman en el suelo a causa de la luz solar que entra por las ventanas son casos de límites proyectados.

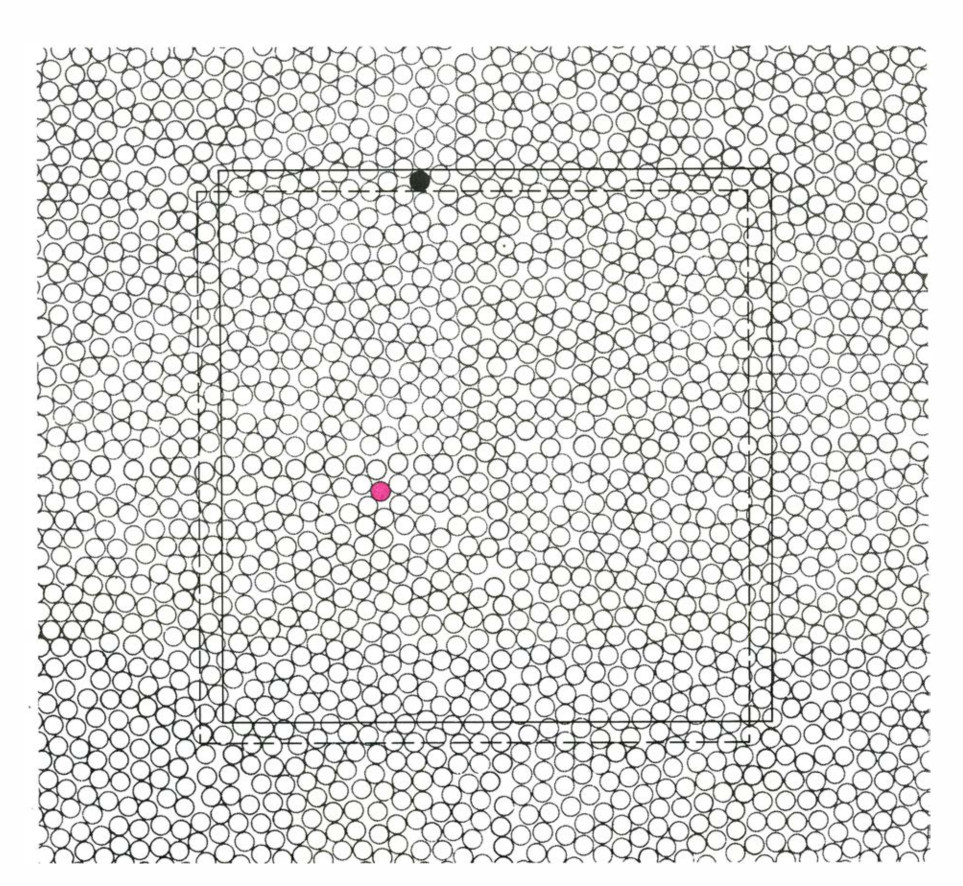




LAS CELULAS RECEPTORAS DE LA RETINA obtienen información al detectar los cambios de luz. Para detectar tales cambios debe existir un desplazamiento relativo y continuo entre la imagen retiniana y la retina; de esta forma el ojo explora una superficie moviéndose continuamente hacia adelante y hacia atrás con unas fluctuaciones que tienen una frecuencia entre 30 y 150 ciclos por segundo. En el test de la izquierda, la frontera entre el disco verde y el anillo rojo se hizo mover de suerte que siguiera las vibraciones oculares. Esto suprimió todo desplazamiento entre las células y la frontera. La frontera en la retina desapareció y entonces el observador veía un disco rojo (*derecha*). El experimento fue realizado por Krauskopf.

tera entre el anillo rojo y el disco verde estuviese estabilizada; es decir, se movía siguiendo las fluctuaciones oculares. De esta forma quedó eliminado todo desplazamiento de la frontera verde-roja con respecto a las células receptoras. Como consecuencia, la frontera desapareció en la retina y el observador sólo veía un gran disco completamente rojo.

La interpretación más obvia de este experimento se basa en que el ojo sólo envía información al cerebro acerca de los cambios de luz que se producen en los contornos; el cerebro considera homogéneas aquellas áreas en que no se observa ningún cambio. En el test de Krauskopf y en la forma usual de observar un disco totalmente rojo, el ojo sólo extrae información de la parte exterior del disco. Al no detectarse ningún cambio dentro de la frontera, el cerebro considera la parte interior del disco como homogénea. Veamos qué supone esto. Los observadores de Krauskopf percibían el rojo en el centro del disco, a pesar de que sobre la correspondiente región de la retina incidía luz verde. Por tanto, si el color de la luz (verde) que incide para formar la imagen del disco que se forma en la retina no tiene nada que ver con el color percibido (rojo), por el mismo razonamiento el color de la luz (rojo) que forma la imagen retiniana del anillo puede no tener ninguna relación con el color rojo que se percibe.



ESTIMULACION DE UNA CELULA RECEPTORA (*punto negro*). La estimulación cambia cuando el límite de la imagen retiniana se mueve a través de ella a raíz del cambio de posición (*rectángulo continuo*) a otra (*rectángulo discreto o de líneas interrumpidas*), a causa de las fluctuaciones oculares. Las células como la que está coloreada sobre las que no se produce movimiento de la frontera de la imagen retiniana no reciben ninguna variación en la estimulación. Como la frontera verde-roja en el test o dispositivo de Krauskopf no se movía a través de ninguna célula, el test parecía rojo.

El test de Krauskopf sugiere que, incluso en condiciones naturales, el color que se percibe de una superficie no depende del color de la luz que de ella nos llega, sino del cambio en la luz que pueda existir en el límite de la superficie. El color percibido de una superficie no puede depender tan sólo del cambio de la luz en el borde de la superficie, porque un cambio de esta naturaleza es estrictamente relativo. El cambio de la luz en el borde de la superficie depende tanto del color del fondo como del color de la superficie.

Si el color percibido dependiese tan sólo del cambio de luz en el límite de la superficie, debería verse la superficie de colores radicalmente distintos al ser observada sobre diferentes fondos. Por el contrario, dos superficies de colores totalmente diferentes podrían parecer del mismo. Por ejemplo, el cambio de intensidad en la luz reflejada por una superficie blanca con respecto a un fondo gris intermedio es el mismo que el que se produce entre un fondo negro y una superficie de un gris intermedio. Por supuesto que una superficie blanca y una gris intermedio no aparecerían como iguales vistas sobre estos fondos respectivos. Estos cambios en el color del fondo pueden originar los llamados efectos de contraste; pero estos efectos no son, con mucho, lo suficientemente importantes como para que la superficie blanca sobre el fondo gris intermedio se asemeje a la superficie de color gris intermedio colocada en un fondo negro. Es un hecho empíricamente destacable que el color que se percibe de una superficie permanece ampliamente invariable, a pesar de que puedan producirse modificaciones en el fondo, lo que origina variaciones en la información proviniente de los límites de las superficies.

De estos límites, el sistema visual extrae sólo información relativa. Esta información es una pequeña parte de la que el sistema visual necesita para percibir el color, una parte que sólo puede ser interpretada en el contexto de la información relativa obtenida de otros muchos límites de superficies. Hay pruebas de que el sistema visual integra las señales de origen nervioso que provienen de distintos bordes. Registrar los cambios en la luz y a continuación integrarlos es matemáticamente equivalente a registrar la luz punto por punto. Registrar sin más los cambios es, sin embargo, más eficaz y requiere menos capacidad de transmisión. Para entender esto, consideremos una situación parecida que podemos ubicar en un almacén de mercancías. Para dibujar una gráfica en la que se refleje cómo varía día a día el precio de una determinada mercancía, cabe registrar el valor que dicha mercancía alcanza cada día, o bien anotar los días en los que se han producido modificaciones en el precio. El segundo método requiere menos capacidad de transmisión y sería preferible si, para obtener la cotización de una determinada mercancía, se necesitara de una llamada telefónica transcontinental. La información así recogida podría ser fácilmente integrada para dar los precios diarios de la mercancía. Creo que en el sistema visual se produce un sistema similar en la extracción e integración de la información. Edwin H. Land, de la Polaroid Corporation, también ha sugerido la posibilidad de este tipo de procesos.

Parece, sin embargo, que el modelo de la extracción e integración de la reflectancia en los límites está muy condicionado en su posible aplicación. El modelo funciona bien cuando los objetos sometidos a examen están uniforme-

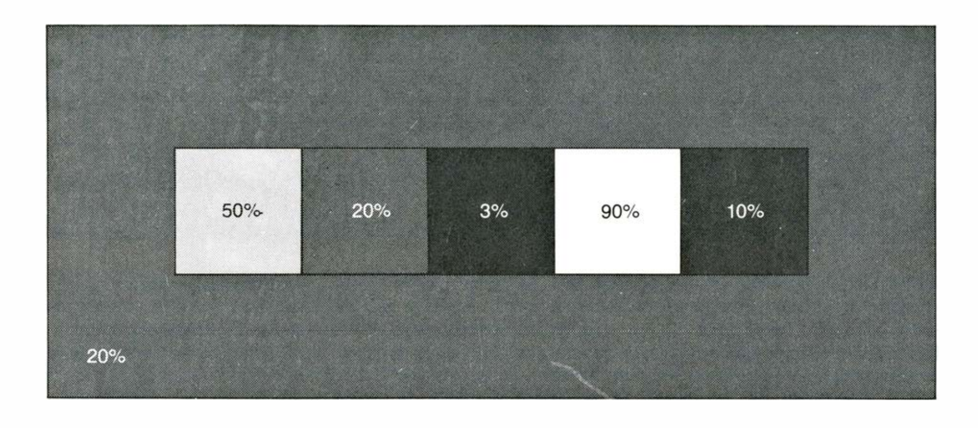
mente iluminados. Incluso muestra su validez cuando la iluminación varía de un modo uniforme sobre todo el campo visual, porque los cambios en la iluminación total no producen ningún efecto en los cambios relativos de la luz en la frontera entre dos tonos de gris. Esto no es más que un caso particular del fenómeno general por el que un cambio de iluminación no modifica la relación de la intensidad de la luz reflejada por superficies adyacentes de reflectancias distintas. Consideremos dos superficies cuyas reflectancias sean, respectivamente, el 25 y el 50 por ciento. La relación de la luz reflejada seguirá siendo de 1:2, sin que influya la intensidad de la iluminación. Se ha demostrado que la señal nerviosa que se origina en el límite de la imagen retiniana permanece constante si la relación de intensidades en el borde o límite permanece invariable.

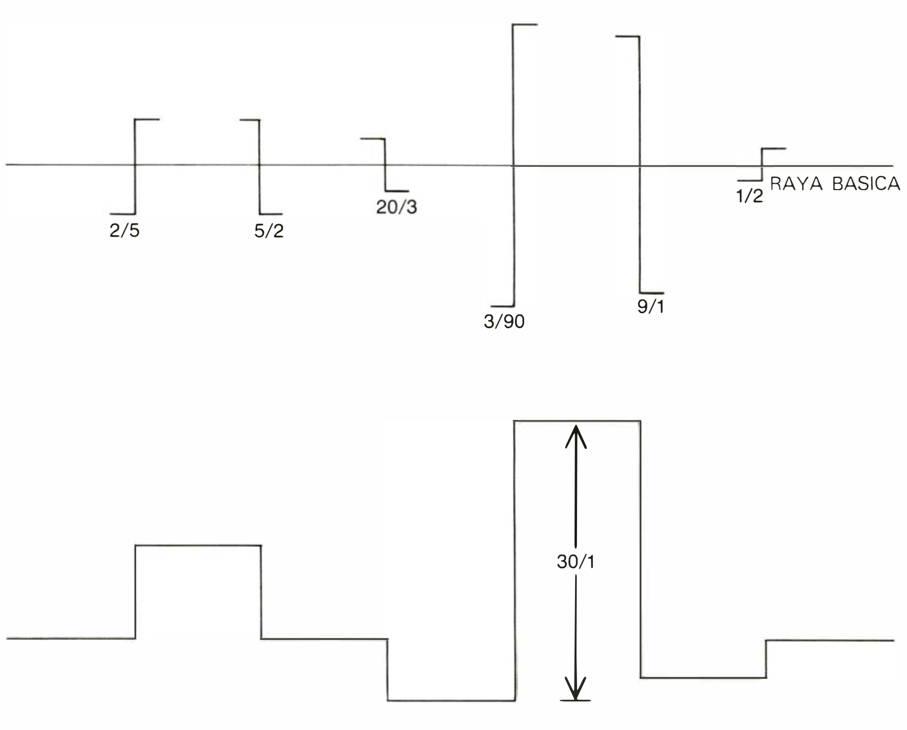
El modelo de la extracción e integración de las reflectancias en los bordes presenta problemas cuando los objetos sometidos a examen no se hallan uniformemente iluminados. Los problemas comienzan cuando aparecen cambios espaciales en la iluminación. Estos límites en la iluminación pueden ser de dos tipos: intrínsecos y proyectados. Los límites intrínsecos tienen su origen en los cambios de planaridad, u orientación espacial, de las superficies. Si se juntan dos paredes del mismo color en una esquina y una de ellas recibe más iluminación que otra, existe un borde de iluminación intrínseco allí donde las paredes se juntan. Tales límites aparecen también en el caso de una escultura realizada con algún material totalmente blanco, donde las irregularidades en su forma se distinguen con claridad porque la intensidad de la iluminación incidente varía según el ángulo que forman la dirección de la luz que ilumina con las distintas superficies que la conforman.

Por el contrario, la frontera de una sombra es un caso de borde o límite proyectado. Como lo es el contorno de la superficie iluminada por un proyector sobre una pantalla, o el de la superficie iluminada de forma trapezoidal que se forma en el suelo de una habitación cuando la luz del sol entra por una ventana rectangular.

La presencia de bordes de iluminación en la imagen retiniana hace impracticable el simple modelo de la extracción e integración de la reflectancia en los límites. Si los cambios de iluminación fueran extraídos e integrados junto con las diferencias de reflectancia, ocurrirían grandes errores en la iluminación percibida. Así, por ejemplo, una superficie blanca en la sombra puede aparecer como una superficie gris. Sin embargo, el modelo se puede perfeccionar para eliminar tales deficiencias. Se impone el que, antes que los límites sean integrados en una imagen de intensidad, se clasifiquen como cambios de iluminación o cambios de reflectancia. Una vez realizada tal clasificación, el sistema visual podría integrar ambos tipos de límites. El resultado sería dos imágenes distintas, una que representara la reflectancia de la superficie y la otra la iluminación de la superficie.

Mi propio trabajo se ha centrado en determinar las condiciones bajo las cuales el sistema visual puede distinguir entre diferencias de reflectancia y diferencias de iluminación y en determinar la manera como podría acometerse tal distinción. He dado por sentado que la distinción de los límites no puede hacerse a nivel de la retina. Ocurre que los cambios espaciales en iluminación y los cambios en reflectancia generan bordes idénticos sobre la retina. Algunos investigadores han sugerido que la retina podría distinguir los tipos de límites en razón de su nitidez, ya que las diferencias de reflectancias tienden a ser muy nítidas y las diferencias en iluminación más graduales. Sin embargo, esta simple explicación no parece ser muy satisfactoria, porque los límites de iluminación (tanto intrínsecos como proyectados) son en muchas ocasiones nítidos, mientras que se dan casos de límites de reflectancia graduales. A pesar de lo cual, el



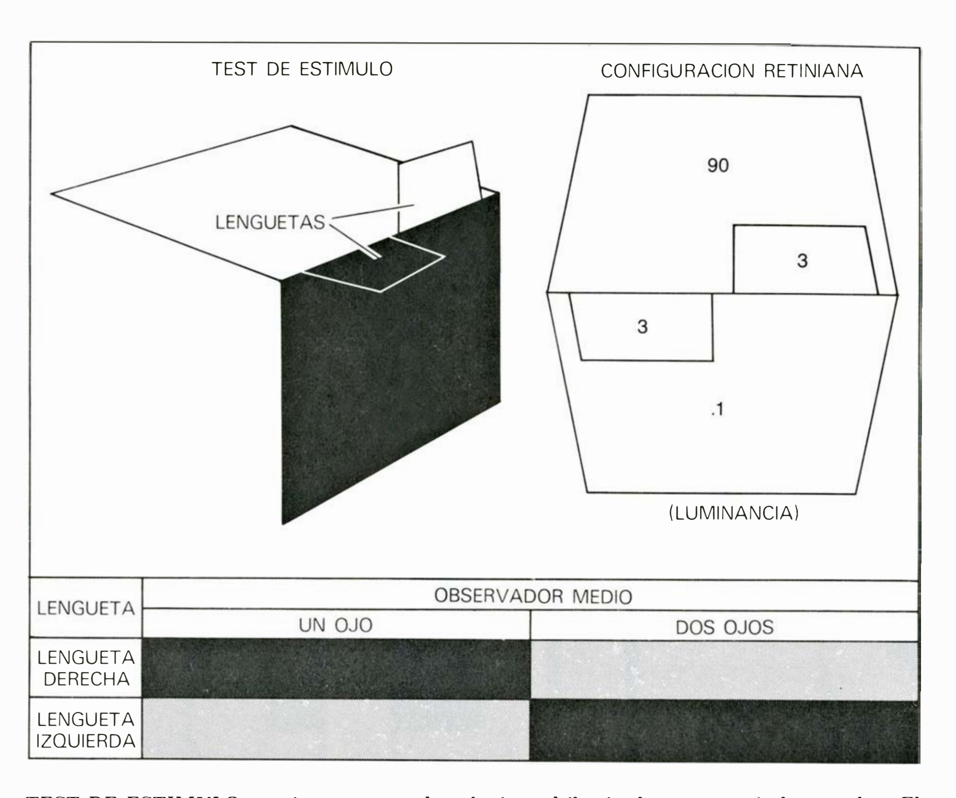


REFLECTANCIA DE UNA SUPERFICIE. Equivale al tanto por ciento de iluminación que refleja la superficie. La parte superior de la ilustración indica la reflectancia de distintas superficies en una escena visual. El ojo sólo responde a los cambios de intensidad en la luz reflejada. La parte central de la ilustración indica el tipo de información que se genera en los límites de la imagen retiniana. Cada línea vertical indica el cambio de intensidad en un límite. El sistema visual integra la información proviniente en todos los límites para reconstruir el patrón de intensidad original (*parte inferior del diagrama*).

sistema visual parece identificar correctamente ambos tipos de límites. Más que clasificarlos sólo en base a su nitidez, el sistema visual los clasifica relacionándolos con todos los límites que aparecen en la imagen retiniana. Esto significa que, además del sistema visual, el sistema nervioso central desempeña también un papel en la clasificación de los límites.

El que el sistema nervioso central esté involucrado en la percepción del color, me lo sugirió un fenómeno muy simple que Irving Rock, de la Universidad de Rutgers, me puso de manifiesto. Observó que, generalmente, la parte más sombreada de una esquina (por ejemplo, donde se cortan dos paredes) parece presentar el mismo grado de gris que la parte iluminada. Pero si miramos por un agujero de suerte que la esquina nos parezca plana, advertiremos que la parte no iluminada se nos muestra de un intenso gris obscuro y, a veces, hasta negra. Al contemplar así la esquina no se perciben sombras, el obscurecimiento se atribuye a la propia superficie.

En esta situación se puede percibir el límite en la esquina como límite de iluminación cuando se le observa correctamente, o como límite de reflectancia cuando se hace aparecer como plano. Ello sugiere que el sistema visual precisa de la percepción del relieve para determinar si es la iluminación o la reflectancia la que está variando en un límite. Más aún, la situación parece poner en cuestión la opinión generalizada de que la iluminación se determina por relaciones de intensidad de regiones adyacentes en la imagen retiniana, con independencia de que estas regiones se perciban o no como si estuvieran en un espacio tridimensional. Para probar concluyentemente que la percepción del relieve (y por tanto el sistema nervioso central) participa en la percepción del color, he tratado de reproducir el fenómeno de la esquina en unas condiciones estrictas de laboratorio.



TEST DE ESTIMULO, consistente en un plano horizontal iluminado y otro vertical en sombra. El sistema o test fue contemplado por los observadores con un ángulo de 45 grados. Como las lengüetas eran trapezoidales, parecían coplanares con los cuadrados grandes cuando el observador miraba con un solo ojo a través de una abertura. Cuando se miraron las lengüetas con los dos ojos, se manifestaron con su verdadera disposición espacial. En ambos casos, el patrón retiniano era el mismo, aunque la luminosidad percibida varió. En esta ilustración, el esquema inferior muestra lo que vieron los observadores. El tono gris percibido resultó depender de la relación entre la intensidad de la lengüeta y la intensidad de toda superficie que pareciera estar situada en el mismo plano que la lengüeta, aun cuando esa superficie no fuera el fondo de la lengüeta en la imagen retiniana. (Ilustración de Ilil Arbel.)

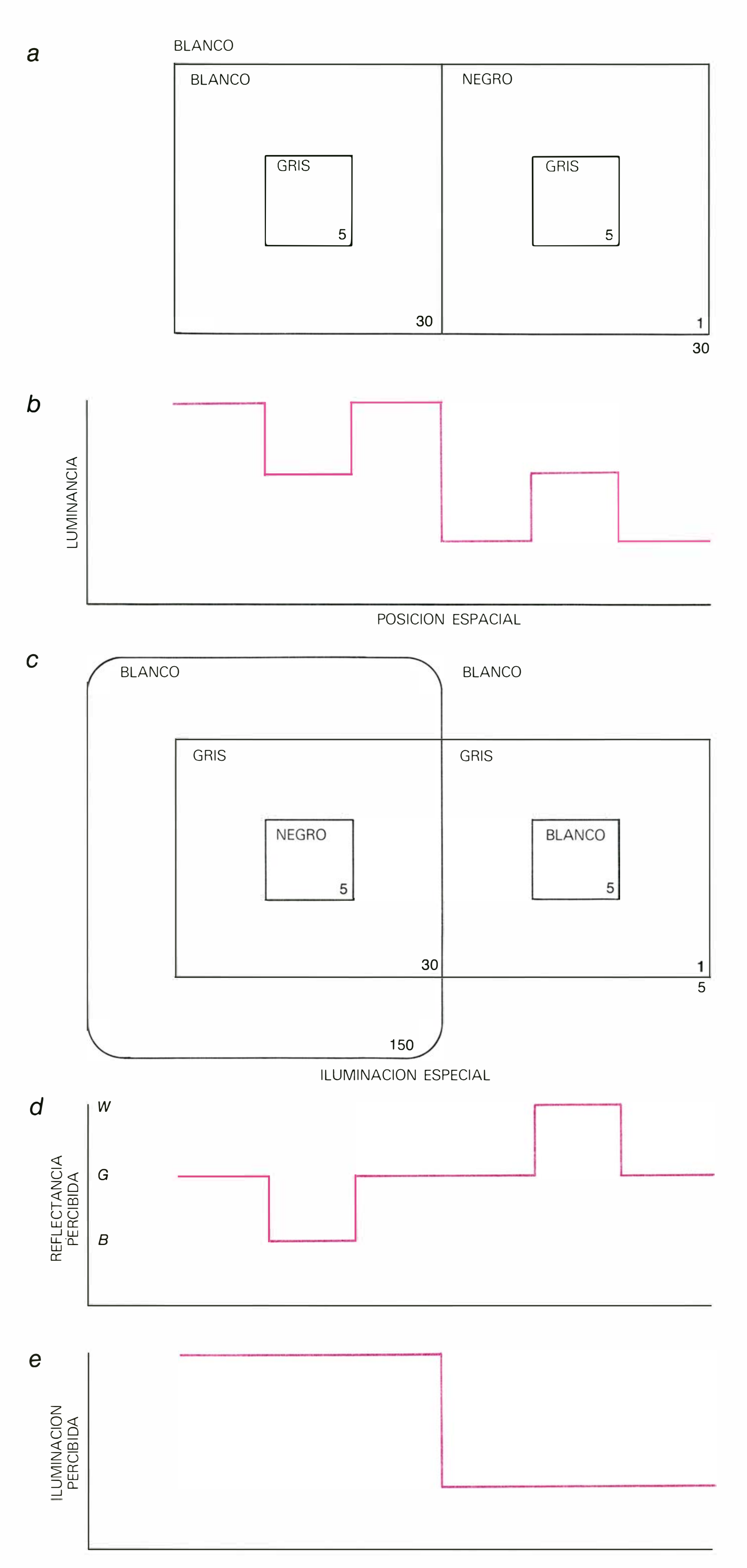
Por medio de un soporte oculto, suspendí en el aire dos superficies que se encontraban formando exteriormente un ángulo recto. Tras las superficies, dispuse un fondo uniforme de intensidad media. Una de las superficies blancas recibía 30 veces más luz que la otra. Los observadores veían el montaje e indicaban la luminosidad aparente de cada superficie eligiendo unos grises en una tabla en la que estaban representados distintos tonos. Un segundo grupo de observadores observó las superficies cuando parecían estar en un mismo plano mirando a través de un orificio con un solo ojo.

Cuando las superficies aparecían planas, la diferencia de intensidad se percibía como una diferencia de reflectancias, que era lo esperado. En otras palabras, el lado iluminado (le daremos un valor 30) parecía blanco, mientras que el sombreado (valor 1) parecía negro. Sin embargo, para mi sorpresa, los observadores vieron exactamente lo mismo cuando ambas superficies aparecían formando un ángulo recto. Fundamentalmente eso era lo que investigadores anteriores habían descubierto en circunstancias similares. En algunos estudios precedentes se había encontrado que la disposición percibida tenía una pequeña influencia sobre la luminosidad captada, pero en la mayoría de las investigaciones, incluida la mía, no se manifestaba tal efecto.

Estaba tan impresionado por la potencia del efecto de la orientación espacial, cuando se observaba en condiciones naturales, que pensé que debía haber algo significativo en el contexto natural de una esquina que había faltado en mi simulación de la situación en el laboratorio. Intenté distintas formas de enriquecer el conjunto que formaba el test del laboratorio. Así coloqué distintos objetos por las cercanías de la esquina para hacer patentes a los observadores las condiciones de iluminación. Pero ningún intento logró modificar los resultados.

Al final conseguí el efecto de la orientación espacial cuando dispuse una superficie negra, cercana a cada una de las superficies blancas que formaban la esquina y coplanar con cada una de ellas. Como en el test inicial, la relación de intensidad de 30:1 entre las dos superficies blancas se percibió como una diferencia de reflectancia cuando la esquina parecía plana. En el caso de observación de superficies que formaban ángulo recto, se percibió la misma relación de intensidades como una diferencia de iluminación. En otras palabras, cuando las superficies aparecían coplanares, la superficie iluminada se mostraba blanca y negra la no iluminada; pero cuando las superficies se observaban tal como estaban dispuestas en la realidad, es decir, formando una esquina, la superficie ensombrecida se vio correctamente como una superficie blanca en la sombra.

Una interpretación razonable de estos resultados consiste en que el sistema visual opera para dar cuenta de los cambios de intensidad en la retina como si provinieran indistintamente de cambios



de iluminación o de cambios de reflectancia. El test inicial con dos superficies pudo verse como si estuviera uniformemente iluminado porque los dos niveles de intensidad (30 y 1) pudieron considerarse como si provinieran de dos tonos de gris, a saber blanco (30) y negro (1). En el test más complicado de cuatro superficies aparecen relaciones de intensidad de 900 a 1. (La superficie blanca iluminada es 900 veces más intensa que la zona negra sin iluminar.) Como el cambio en la intensidad de tonos de gris que puede resultar de cambios de reflectancia está normalmente limitado a 30:1, en el test de las cuatro superficies se precisa la percepción de alguna variación de la iluminación. De la organización de la escena, y en particular de la distribución tridimensional percibida, depende qué límites vayan a tomarse como representativos de cambios de iluminación y cuáles como representantes de los cambios de reflectancia.

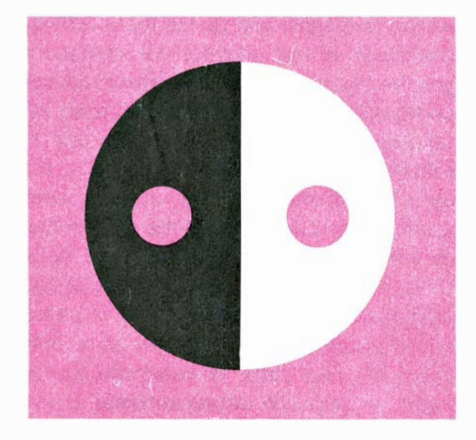
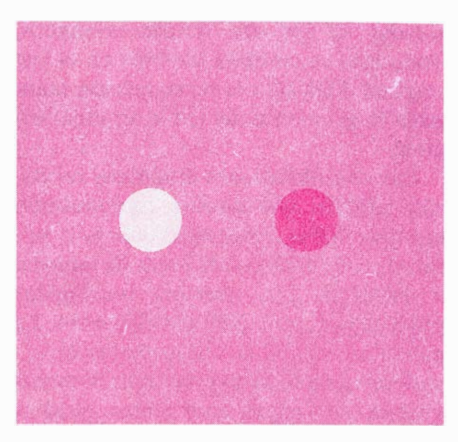
A continuación modifiqué el sistema de las cuatro superficies para descartar la posibilidad de que los resultados estuvieran originados por algún efecto de contraste entre áreas adyacentes en la imagen retiniana bidimensional. El plano horizontal del test modificado consistió en un gran cuadrado blanco, unido a una pequeña lengüeta negra de forma trapezoidal que se extendía hacia el observador [*véase la ilustración de la página anterior*]. El plano vertical estaba formado por un gran cuadrado blanco unido a una pequeña lengüeta blanca trapezoidal que se extendía hacia arriba. En otras palabras, las lengüetas se extendían desde la esquina formada por los dos grandes cuadrados; así, cada lengüeta aparece con tres lados contra un fondo que no está en el mismo plano. Como en el caso original del test de las cuatro superficies, el plano horizontal recibía 30 veces más luz que el vertical. Las lengüetas son trapezoidales para crear la ilusión de una percepción en relieve.

TEST DEL CONTRASTE de luminosidad (*a*). Da lugar a un patrón de intensidad (*b*) que está formado por límites de reflectancia. Los dos cuadrados grises son percibidos como si tuvieran casi el mismo tono de gris intermedio. Un patrón de intensidad idéntico puede provenir de diferencias de iluminación y de un test alterado (*c*). Cuando los observadores miraron el test pudieron ver claramente que un haz de luz rectangular iluminaba la mitad del test (*e*); por tanto, el cuadrado iluminado fue percibido correctamente como negro y el no iluminado como blanco (*d*). Cuando se contempló el test a través de una abertura que sólo permitía a los observadores ver el rectángulo gris, el sistema visual no pudo reconocer que el rectángulo gris no estaba iluminado uniformemente, con lo que parecía el test original de reflectancia.

Vistas con ambos ojos, las lengüetas aparecen correctamente como trapezoides que están en sus respectivos planos. En este caso, la lengüeta vertical parece casi blanca y la horizontal casi negra. Pero observado el test con un solo ojo a través de un orificio, cada lengüeta se nos antoja un pequeño cuadrado contenido en el mismo plano que el cuadrado grande, que le rodea por tres lados. En este caso, los colores percibidos están invertidos, la lengüeta vertical se muestra negra y la horizontal blanca.

Pero en ambos casos la imagen retiniana es la misma, cada lengüeta se ve contra un fondo cuadrado que la rodea por tres lados. Ello significa que la relación entre el objeto y su fondo en la imagen retiniana es irrelevante por lo que respecta a la percepción del tono de gris del objeto. El tono de gris resulta de la relación entre la intensidad del objeto y la intensidad de cualquier superficie que parezca que se encuentra en el mismo plano, aun cuando la superficie coplanar no proporcione el fondo del objeto en la imagen retiniana. Por tanto, cuando observamos el sistema con los dos ojos, la lengüeta blanca vertical ensombrecida aparece cual si fuera coplanar con el fondo vertical negro ensombrecido y, por tanto, la lengüeta se ve blanca. Además, la lengüeta negra horizontal, que está iluminada, se muestra coplanar con el fondo blanco horizontal que está iluminado, y, por tanto, la lengüeta se ve negra. Cuando se observa el test o sistema con un solo ojo, la lengüeta vertical blanca en sombra parece coplanar con el fondo blanco horizontal iluminado, y, por tanto, la lengüeta se ve negra. Más aún, la lengüeta negra horizontal iluminada aparece coplanar con el fondo negro vertical ensombrecido; por ello, la lengüeta se ve blanca.

El modelo de la extracción, clasificación e integración de los límites fue sometido a una prueba más directa en una serie de experimentos diseñados para ensayar qué verían varios observadores si un límite de reflectancia se percibiera erróneamente como un límite de iluminación, o viceversa. Los experimentos con la esquina que acabo de describir probaron que un cambio en el tipo de límite percibido se traducía en un cambio de la iluminación percibida de la región unida al límite que se ha observado con alteraciones. Si la luminosidad percibida depende de la integración por el sistema visual de una serie de límites alejados espacialmente, entonces se manifestarán otros errores en la percepción. Por ejemplo, si en un test muy complejo hacemos que un límite de reflectancia sencillo aparezca como un límite de iluminación, entonces, en ciertas circunstancias ello debería incidir sobre la luminosidad percibida de las regiones que se encuentren lejos del límite alterado con tanta eficacia como pueda tener sobre las regiones unidas al límite alterado.

LOS OBSERVADORES vieron el test de la izquierda en unas condiciones en que no se podían percibir los fondos blancos y negros inmediatos porque la frontera entre ellos, y también la frontera entre cada uno de ellos y el fondo rojo común a los dos, se movía adelante y atrás, siguiendo las fluctuaciones oculares. Los círculos rojos parecían entonces estar dispuestos sobre un fondo rojo homogéneo (*derecha*). La frontera del círculo de la derecha todavía contenía la información de que el círculo era más obscuro que su fondo inmediato: así, apareció de un rojo muy obscuro. Con el círculo de la izquierda sucedía lo mismo: la frontera contenía la información de que el círculo era menos obscuro que su fondo inmediato; por tanto, aparecía de un rojo muy claro. Este experimento, realizado por Alfred Yarbus, de la Academia de Ciencias de la Unión Soviética, pone de manifiesto que un cambio introducido en la información contenida en un límite afecta a zonas que están espacialmente alejadas del límite alterado.

En la Universidad estatal de Nueva York en Stony Brook, Stanley Delman y yo diseñamos un experimento de substitución de límite. Estudiamos un test familiar, conocido como contraste simultáneo de luminosidad, que suele describirse en los libros de psicología. Se sitúan dos cuadrados de un gris medio sobre fondos negros y blancos adyacentes, respectivamente. Sucede que, por alguna razón, el cuadrado gris, que está colocado sobre fondo blanco, parece ligeramente más obscuro que el que está situado sobre fondo negro. Pero este efecto no era el tema de nuestro experimento. Queríamos reproducir la misma disposición de luz reflejada que el ojo recibe del test cuando la frontera entre los fondos blanco y negro se percibía como la frontera entre niveles intensos y débiles de iluminación.

Con este fin colocamos en una pared una pieza grande rectangular de papel gris intermedio que sirviera de fondo para ambos cuadrados. La diferencia en las intensidades de ambos fondos se creó proyectando un haz luminoso que tenía un límite nítido a lo largo de la mitad del rectángulo gris. Una parte del haz iluminaba también la pared por tres lados del rectángulo. Se fijó la intensidad del haz de forma que la relación de intensidades entre la zona iluminada y la zona en sombra del rectángulo gris fuera de 30:1, que es la misma relación de intensidades entre un papel negro y otro blanco, tal como ocurría en el test original. Teníamos que añadir luego los dos cuadrados. No podían tener el mismo tono de gris (es decir, la misma reflectancia), porque hubieran reflejado distinta intensidad de luz hacia el sistema visual, por su iluminación desigual.

En el test original de contraste de luminosidad que tratábamos de reproducir, los dos cuadrados reflejaban la misma cantidad de luz. También es cierto que ambos cuadrados tenían la misma reflectancia en el test original. Tampoco esto importaba, pues no nos proponíamos reproducir todas las características del test original, sino recabar qué información recibía el sistema visual: la disposición de las intensidades luminosas que llegaban al ojo. Para lograrlo, el cuadrado de la parte no iluminada debía poseer una reflectancia 30 veces mayor que el de la parte iluminada, que iba a recibir 30 veces más luz. Colocamos, pues, un cuadrado blanco sobre la parte no iluminada y otro negro sobre la iluminada.

De esta forma, pensábamos que habíamos reproducido la relación de intensidades del test original de un modo lógico. Era el momento de comprobar nuestros resultados fotométrica y empíricamente. Las mediciones fotométricas pusieron de manifiesto que los cuadrados reflejaban iguales cantidades de luz, que el fondo reflejaba 30 veces más luz en una parte que en la otra y que, en escala logarítmica, las intensidades de

los cuadrados eran la media de las intensidades de los dos fondos contiguos. Todo esto se cumplía también en el test original. Más aún, la reproducción mantenía la geometría del test o sistema original.

Las consideraciones empíricas revelaron también que nuestra reproducción era muy fiable. Los observadores la miraron, bajo unas condiciones de control, a través de una abertura rectangular practicada en una pantalla de cartón. La pantalla limitaba la visión de los observadores de suerte que las únicas partes visibles del sistema eran ambos cuadrados y sus fondos inmediatos: la mitad iluminada y la mitad en sombra del rectángulo gris. En estas condiciones, el test parecía idéntico al original. Los fondos contiguos se mostraban uno blanco y el otro negro, aunque en realidad se trataba de un gris intermedio que había sido diferentemente iluminado. El límite de iluminación entre los fondos se percibía como un límite de reflectancia y los cuadrados parecían ser de un gris medio, uno un poco más obscuro que otro.

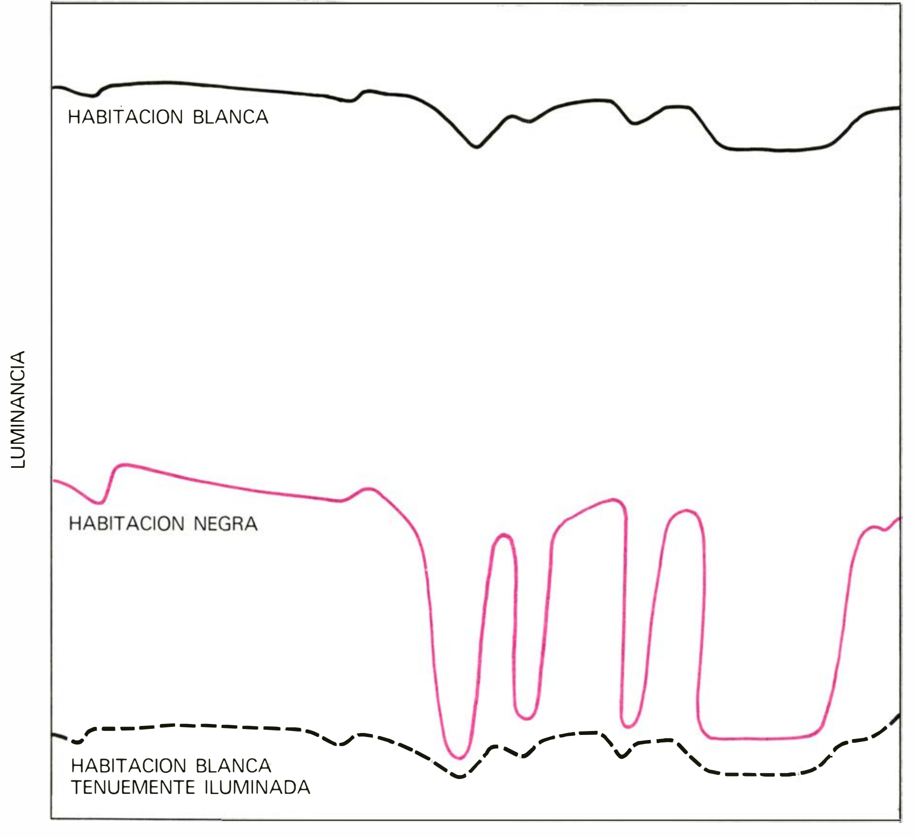
Todo esto sirvió para asegurarnos de que habíamos reproducido correctamente la disposición de las luces reflejadas según el test original. La verdadera prueba llegó cuando los observadores contemplaron el test o dispositivo sin estar condicionados en su visión por el rectángulo practicado en la pantalla. En estas circunstancias, los observadores pudieron percibir claramente que un haz de luz estaba iluminando la mitad de la pieza rectangular de papel, ya que el haz luminoso incidía también sobre la pared sobre la que se había colocado el rectángulo. En este caso, el cuadrado iluminado se percibió correctamente como negro, mientras que el no iluminado se vio blanco, también correctamente.

La luminosidad percibida de cada cuadrado cambió totalmente del test apantallado al no apantallado, a pesar de que la cantidad de luz reflejada por cada cuadrado era exactamente la misma que en el caso anterior y que, por tanto, la señal nerviosa generada en los límites de la imagen retiniana tenía que ser idéntica. Esto significa que la percepción del color no es sólo función de la cantidad de luz que refleja una superficie, de la relación de intensidades de las luces reflejadas por el fondo y la superficie, o de la señal nerviosa generada en la imagen retiniana por el límite de una superficie.

Todo indica, pues, que en el sistema visual ocurre un proceso mucho más relativista. La frontera entre cada cuadrado y su fondo proporciona solamente la relación de luz reflejada por estas dos superficies. Si cada cuadrado se compara con el otro, debemos tener en cuenta la relación de intensidades de los fondos. Si los fondos contiguos se perciben con reflectancias radicalmente distintas, entonces esta percepción, junto con los límites entre los cuadrados y los fondos, sugiere que los cuadrados tienen aproximadamente la misma luminosidad. Por otra parte, si los fondos se perciben difiriendo completamente en la iluminación, entonces los cuadrados parecen estar situados sobre el mismo fondo de reflectancia, y, por tanto, parecen diferir de un modo absoluto en la luminosidad.

Llegados a este punto, consideremos cómo se explican los anteriores resultados con el modelo de la clasificación de los límites. En el caso de la observación apantallada, en que el test parecía el original de contraste de luminosidad, los cambios en los límites podían asumirse e integrarse para obtener un perfil de intensidades en el test. En el caso donde la observación se efectuaba sin ningún tipo de apantallamiento, se hubiera generado el mismo perfil de no ser que los límites no hubieran sido clasificados como límites de iluminación o de reflectancia. Sin embargo, si los límites son clasificados antes de la integración y si la integración se realiza sólo dentro de cada tipo de límites, entonces se generan dos perfiles separados. Aquí, el perfil de reflectancia del test apantallado debería ser diferente del no apantallado. La diferencia estriba en que la discontinuidad central, que representa la relación entre los fondos adyacentes, aparece sólo en el perfil de reflectancias del test apantallado. Esto significa que, en el test no apantallado, los cuadrados aparecen con reflectancias claramente diferentes (de hecho como blanco y negro) en lugar de mostrarse como dos tonalidades de gris intermedio. Por supuesto que el límite entre los fondos contiguos que falta en el perfil de reflectancia se manifiesta en el perfil de iluminación. Así pues, el perfil de iluminación mostraría una región de gran iluminación al lado de otra de poca iluminación. En el test apantallado, el perfil de iluminación sería sencillamente uniforme.

Alfred Yarbus, de la Academia de



ILUMINACIONES INDIRECTAS, que parecen intervenir en el proceso a través del cual el sistema visual determina la intensidad de gris de una superficie. La curva superior presenta el perfil de la intensidad de una habitación amueblada que estaba pintada totalmente de blanco; la curva inferior, la misma habitación tenuemente iluminada; la curva central, una habitación idéntica pintada totalmente de negro. En las habitaciones blancas, las iluminaciones indirectas difuminaron la nitidez de los límites de intensidad que resultaban de la iluminación directa; así los perfiles de intensidad (*superior e inferior*) son comparativamente más difusos. En la habitación negra, había muy poca iluminación indirecta, con lo que el perfil (*en el centro*) presentaba una variación con saltos arriba y abajo característicos de los límites generados por iluminación directa. El tono de gris percibido no dependía de la intensidad de la luz reflejada, porque la habitación negra reflejaba más luz que la blanca, tenuemente iluminada.

Ciencias de la Unión Soviética, realizó un experimento similar. Usó una técnica según la cual la imagen de la retina está estabilizada aunque el ojo se mueva hacia adelante y hacia atrás. Dispuso una región negra junto a una blanca sobre un fondo rojo [*véase la ilustración de la página 67*]. Se colocaron sendos círculos de un rojo igual al rojo del fondo en el centro de la zona blanca y en el de la negra. En condiciones normales de visión, el círculo de la región blanca aparecía un poco más obscuro que el de la negra. Yarbus alteró las condiciones eliminando las fronteras de percepción entre la zona blanca y la negra y entre estas regiones y el fondo rojo. Consiguió esto por medio de un dispositivo que hacía que las fronteras físicas se movieran junto con el ojo, como había hecho Krauskopf con el disco verde y el anillo rojo. Como el ojo sólo registra cambios en la intensidad de la luz cuando existe un desplazamiento relativo constante entre la imagen retiniana y la retina, no pudo detectar la presencia de las regiones blanca y negra. En consecuencia, los círculos parecían estar dispuestos sobre un fondo rojo homogéneo. Yarbus no publicó sus resultados de una forma cuantitativa, pero indicó que un disco (el que objetivamente se encuentra sobre la región blanca) aparecía muy obscuro, mientras que el otro se veía muy claro.

Los resultados de Yarbus fueron esencialmente los mismos que los míos, aunque existían diferencias significativas en cuanto a nuestros métodos que son dignas de exploración. En el experimento de Yarbus, y en el caso de mi test sin apantallar, los círculos en un caso y los cuadrados en otro fueron hechos aparecer como si tuvieran un fondo común eliminando del perfil de reflectancia el límite que dividía el fondo. Por lo que respecta a la luminosidad percibida, los efectos son los mismos. Lo que resulta fascinante es que Yarbus eliminó el límite en el proceso de extracción, mientras que yo lo eliminé en el proceso de clasificación. Dicho con otras palabras, en el experimento de Yarbus, el ojo nunca detecta el límite, mientras que en el caso de mi test no apantallado el ojo lo detecta y lo clasifica correctamente como un límite de iluminación. En el experimento de Yarbus, no existe razón para creer que los observadores veían los círculos como si estuvieran diferentemente iluminados.

El experimento de Yarbus proporciona una experiencia directa de la validez del concepto de extracción e integración de los límites. No cambiaba el límite de cada círculo, sino que las fronteras que los dividían del fondo desaparecieron. Como en el caso del disco verde y del anillo rojo de Krauskopf, la región limitada por la frontera que se evanescía, parecía cambiar de color. En este caso, las regiones blanca y negra del fondo desaparecieron dejando tan sólo un único fondo coloreado de un modo uniforme. Sin embargo, el trabajo de Yarbus revela un fenómeno adicional: los círculos parecían cambiar de color aunque sus fronteras permanecían constantemente visibles. La conclusión obvia es que la información en un límite (en este caso, el límite entre un círculo y su fondo inmediato) se integra de alguna forma junto con la información extraída de un límite remoto (en este caso la frontera entre el fondo inmediato y el fondo que le rodea) antes de que el color de las regiones se perciba, por fin.

Conviene que examinemos ahora por qué el sistema visual determina los colores de las superficies comparando las intensidades de la luz reflejada. Como apunté, esta comparación implica extraer información de los límites o contornos. Lo cual sugiere que es más útil considerar la imagen retiniana como un conjunto de límites que como un mosaico de manchas de color, que hubiera sido el camino de las explicaciones tradicionales de la visión del color. Puesto que los límites son de dos tipos, la imagen retiniana es en realidad una imagen dual.

La retina es una superficie fotosensitiva preparada para registrar gradientes en la excitación óptica. Esencialmente existen dos tipos de cambios en el mundo físico que se traducen en gradientes de estimulación en la retina: las variaciones de reflectancia y de iluminación de las superficies. Cada tipo de variación es capaz de inducir por si solo un patrón complejo de la estimulación en la retina. De una gran extensión de empapelado iluminada uniformemente, la retina recibe exclusivamente un patrón que proviene de los cambios de reflectancia en la superficie. Por el contrario, de las sábanas blancas arrugadas de una cama sin hacer o de un paisaje cubierto por la nieve, la retina recibe un patrón de estimulación que proviene exclusivamente de las variaciones de iluminación. En general, estos dos tipos de límites actúan simultáneamente. Es como si dos patrones distintos estuvieran superpuestos el uno sobre el otro en la retina. Al sistema visual cumple desenredar estos patrones.

La dualidad de la imagen retiniana implica que cada punto del sistema visual contiene al menos dos valores: un valor de reflectancia y un valor de iluminación. La percepción de la transparencia conlleva un fenómeno similar en el que aparecen dos valores, ya que una superficie de un color se ve detrás de una superficie de otro color. El mismo fenómeno sucede en las reflexiones. Por ejemplo, al mirar a través de una ventana, el observador ve simultáneamente dos escenas: la escena real que sucede detrás de la ventana y otra que se refleja en la parte anterior de la ventana. Así pues, pueden verse dos colores en el mismo punto.

Permítaseme relatar una experiencia que tuve recientemente. Sobre el salpicadero de mi coche se encontraba un libro de pastas rojas, de suerte que yo podía observar un reflejo rojo cuando miraba por el parabrisas. Me sorprendió observar que los objetos lejanos conservaban su propio color cuando los observaba a través del reflejo rojo. Incluso los objetos verdes eran vistos como tales al observarlos a través de un reflejo rojo. El fenómeno me llamó la atención. Sabía, por supuesto, que cuando se mezclan aisladamente las luces roja y verde dan lugar al amarillo. Pero cuando levanté la mano para eliminar el resto de la escena y observé una mancha de rojo y verde, que pasaba a través de las pequeñas aberturas que se formaban entre mis dedos, vi amarillo. Las aberturas de mi mano habían impuesto los mismos contornos a las dos luces, la roja y la verde; por tanto, se mezclaban para formar amarillo. Sin embargo, en la situación inicial la frontera del reflejo rojo no coincidía con la del objeto verde. En consecuencia, las luces roja y verde aparecían como elementos de imágenes diferentes y no como amarillo. Se desprende, pues, que, cuando se habla de mezcla de colores, podría hablarse de mezcla de límites (cambios de luz) mejor que de mezcla de luz propiamente dicha.

Estas consideraciones indican que la retina no actúa como una fotocélula, midiendo la intensidad y el color de cada punto de una escena. Los dispositivos diseñados por el hombre para medir están proyectados para responder sólo a una estimulación física en un momento dado. Por ejemplo, si un voltímetro es sensible a la impedancia, o todavía peor, a la temperatura, se considera a este medidor como defectuoso. El proponer estos dispositivos como modelos del sistema sensorial humano puede ser una grave equivocación. Contrariamente a los dispositivos de medida hechos por el hombre, el sistema visual humano parece estar capacitado para procesar simultáneamente gran cantidad de variables. Las explicaciones de la percepción del color, que se basaban en mecanismos de medición inadecuados, sostienen que el ojo humano gana en información

sobre la reflectancia de una superficie a base de sacrificar información acerca de la iluminación. Estas explicaciones limitan el sistema visual humano para que funcione sólo en situaciones simples, cuando en realidad puede procesar situaciones complejas extrayendo información de los cambios de reflectancia e iluminación.

Hasta aquí hemos examinado cómo actúa el sistema visual humano con las imágenes de la retina compuestas de límites de iluminación y límites de reflectancia. Es igualmente viable el estudio de la percepción de imágenes que sólo presentan un tipo de límites. Deseaba saber cómo se verían las cosas si faltaran todos los límites de reflectancia de una determinada situación. Con Alan Jacobson, construí dos habitaciones en miniatura que presentaban exclusivamente límites de iluminación. Amueblamos cada cuarto de forma idéntica con un conjunto de objetos de distintas configuraciones y tamaños. Pintamos una de las habitaciones, con todos sus componentes, de blanco mate (no brillante) y la otra de negro mate. Los observadores veían cada habitación a través de una abertura practicada en una de sus paredes, que les impedía ver la lámpara que iluminaba la habitación.

Conviene tener presente que cada una de estas habitaciones proyecta en la retina del observador un patrón de estimulación no uniforme y complejo. En las habitaciones hay límites nítidos en las esquinas, y difusos en las paredes y en las superficies curvas. Algunos de estos límites resultan de sombras proyectadas, pero todos ellos son límites de iluminación, ya que eliminamos los límites de reflectancia cubriendo la escena con una pintura uniformemente reflectante.

A la hora de presentar la habitación al examen de los observadores nos apremiaban multitud de interrogantes. Nos preguntábamos si los límites existentes en las habitaciones se percibirían como cambios de reflectancia, según hubieran predicho las teorías vigentes en la percepción, o como cambios de iluminación, según eran ciertamente. En otras palabras, ¿se observaría correctamente cada habitación con un único tono de gris? Si así fuera, ¿cuál sería ese tono? ¿Parecería blanca la habitación blanca y negra la habitación negra?

Resultó que 22 de nuestros 24 observadores percibieron las habitaciones uniformemente iluminadas. Todos vieron la habitación blanca como blanca. En cuanto a la habitación negra, todos la percibieron como si fuera de un único tono gris, aunque el tono captado variaba de observador a observador. Los tonos oscilaron del negro al gris medio y, en promedio, daban un gris obscuro.

El modelo de extracción, clasificación e integración de los límites parecía indicar que en ambas habitaciones los límites de iluminación serían percibidos como tales y, por tanto, la reflectancia de las habitaciones aparecería uniforme. Sin embargo, el modelo no predice el que los observadores pudieron percibir, aunque de forma poco precisa, el tono correcto del gris. ¿Qué información hace posible dicha identificación? Una hipótesis plausible se funda en que la identificación se hace sobre la base de la intensidad de la luz. En nuestro experimento, la intensidad promediada de la luz reflejada por la habitación negra era menor que la reflejada por la habitación blanca. Pero modificando las condiciones del experimento logramos demostrar la falsedad de la hipótesis. Redujimos la intensidad de la lámpara que iluminaba la habitación blanca y elevamos la que iluminaba a la habitación negra hasta el extremo de que, en la nueva situación, en la habitación negra cada punto reflejaba más luz que su homólogo de la blanca. Se llegó a los mismos resultados: la habitación blanca aparecía blanca y la negra, gris obscura.

Tenemos ya una hipótesis prometedora acerca de cómo el sistema visual determina el tono de gris de las habitaciones, aunque todavía no poseemos una explicación completa. (John Robinson me ayudó a desarrollar esta idea.) En mi opinión, el factor operativo es el efecto de la iluminación indirecta en el patrón de la imagen retiniana. En nuestras habitaciones, cada punto recibía iluminación de dos fuentes: directamente de la lámpara e indirectamente de la luz que reflejaban otras superficies de la habitación. En la habitación negra, con una reflectancia de un 3 por ciento, apenas existía iluminación indirecta. La luz directa era, en casi su totalidad, la responsable de las iluminaciones de las distintas superficies. La luz directa genera límites nítidos y, por tanto, el perfil de intensidad de la habitación negra presentaba variaciones muy bruscas. Aunque la iluminación directa pudiera parecer la misma para las dos habitaciones, existía mucha más iluminación indirecta en la habitación blanca, ya que su reflectancia era alrededor del 90 por ciento. Una abundancia tal de luz indirecta tenía el efecto de difuminar los límites nítidos producto de la iluminación directa; el perfil de intensidades de la habitación blanca presentaba, pues, cambios graduales.

Cuando se redujo la iluminación en la habitación blanca, la forma del perfil de intensidades no cambió, tan sólo sufrió un descenso uniforme en intensidad. Si la habitación blanca se hubiera pintado de un tono más obscuro de gris, la forma de su perfil se habría modificado. Nuestro trabajo pone de manifiesto que las habitaciones obscuras tienen gradientes más nítidos que las habitaciones blancas. De alguna manera, el sistema visual descifra el tono de gris de una habitación a partir de la forma que tiene el patrón de intensidades de la habitación. En este área nuestra investigación está sólo en sus comienzos.

En unión con mis colaboradores he comenzado a extender el análisis sobre los grises a los colores cromáticos. En un experimento comparamos las observaciones de personas que veían una habitación blanca iluminada con luz azul con las observaciones cuando se iluminaba con luz blanca una habitación azul. Todos los observadores pudieron discernir si el azulado provenía de las superficies o de la iluminación, a pesar de que la luz reflejada en las dos habitaciones pudiera ser idéntica. Este resultado prueba claramente que la percepción de la iluminación desempeña también un papel importante en la percepción del color de una superficie cromática.

Si la percepción del color de una superficie se puede explicar o no sin tener en cuenta la percepción de la iluminación, se ha venido debatiendo desde finales del siglo XIX, cuando Hermann Von Helmholtz propuso que el color de una superficie sólo se podría conocer una vez que se hubiera determinado el color de la iluminación. Sin embargo, Helmholtz fue incapaz de explicar cómo se podría determinar el color de la iluminación; la mayoría de las explicaciones subsiguientes sobre la percepción del color tampoco pudieron explicar la iluminación percibida. Los experimentos que he descrito sugieren que Helmholtz andaba en lo cierto al destacar la importancia de la iluminación percibida, pero falló al relacionarla con el color que se percibe de una superficie. Había supuesto erróneamente que el ojo mide la luz. El presente trabajo demuestra que el ojo compara la luz por medio de la extracción de información en los límites; así, el sistema visual obtiene información acerca de la iluminación exactamente de la misma forma que lo obtiene del color de las superficies que hay en una escena. La percepción de la iluminación y la percepción del color de las superficies son procesos paralelos en el sistema visual que están relacionados con la descomposición de la imagen retiniana en dos patrones distintos de iluminación y del color de las superficies.

Canales iónicos en la membrana de la célula nerviosa

El impulso nervioso se genera por el flujo de iones sodio y potasio a través de canales moleculares incrustados en la membrana. El funcionamiento de estos canales iónicos está siendo en la actualidad intensamente estudiado

Richard D. Keynes

La unidad básica de información transmitida de un punto a otro del sistema nervioso de un animal es un potencial de acción único o impulso nervioso. A diferencia de los puntos y rayas en la clave Morse, la forma de cada impulso nervioso es constante, con una amplitud de unos 100 milivolt y una duración fija de alrededor de un milisegundo; el contenido de la información depende simplemente de la frecuencia de los impulsos transmitidos por una determinada fibra nerviosa y de las conexiones que la fibra establece en cada extremo.

La línea de transmisión del potencial de acción es el axón o fibra nerviosa, una prolongación alargada de forma cilíndrica que emerge de cada célula nerviosa periférica. El mecanismo de transmisión en el axón es comparable en cierto modo al que tiene lugar en un cable eléctrico aislado, dada la similitud estructural entre ambos. No obstante, si el núcleo conductor y el aislante de un cable tuvieran las resistencias eléctricas típicas de la fibra nerviosa, la señal se atenuaría rápidamente con la distancia. El potencial de acción, sin embargo, se regenera y mantiene una amplitud constante en su recorrido a lo largo de la fibra nerviosa. Estos hechos se conocen desde hace algún tiempo, pero los detalles moleculares de los mecanismos responsables de los mismos están siendo actualmente descifrados.

Los animales invertebrados dotados de una capacidad de respuesta mecánica rápida ante situaciones de peligro (podrían citarse como ejemplos el coletazo de una langosta, la retirada de una lombriz a su escondrijo y la huida a propulsión de un calamar) han desarrollado axones gigantes para garantizar una conducción rápida de los impulsos que activan la contracción muscular. Para los electrofisiólogos, el axón gigante por excelencia es el axón del calamar. Se trata de una fibra nerviosa de fácil disección, con un diámetro de más de medio milímetro, en la que se pueden implantar complejos sistemas de electrodos para mediciones eléctricas y cuya composición iónica puede modificarse a voluntad del investigador.

La evolución del sistema nervioso ha seguido un curso diferente en los animales vertebrados, donde la velocidad de conducción del impulso nervioso en muchas de las fibras aumenta notablemente debido a la presencia de una capa aislante de material lipídico, la mielina. La existencia de esta capa aislante permite un salto rápido del impulso nervioso desde un nodo de Ranvier al próximo. En cada nodo, la vaina de mielina se interrumpe un corto trecho. La mielinización es una modificación puramente estructural de las fibras y la membrana excitable del axón se comporta de la misma forma en los nodos de una fibra mielínica, donde está expuesta, que en las fibras amielínicas.

La mielinización confiere una notable ventaja evolutiva a los vertebrados. La velocidad alcanzada en una fibra nerviosa amielínica de un invertebrado puede conseguirse con un diámetro 25 veces menor en las fibras mielínicas de los vertebrados. Esta reducción hace, sin embargo, que las fibras mielínicas sean más difíciles de aislar por disección y constituyan un material menos adecuado que las fibras gigantes amielínicas para estudiar las propiedades de la membrana excitable.

Los fenómenos básicos que tienen lugar en la generación del potencial de acción fueron dilucidados hace 25 años por Alan L. Hodgkin y Andrew F. Huxley, que investigaban el axón gigante de calamar en los laboratorios de la Asociación de Biología Marina de Gran Bretaña en Plymouth (contribución por la que compartieron el premio Nobel en 1963). Hodgkin y Huxley mostraron que la excitabilidad eléctrica de la membrana de la célula nerviosa depende de la existencia de un sistema de permeabilidad iónica sensible a los cambios de voltaje. La célula puede utilizar la energía almacenada en forma de gradientes iónicos de concentración; estos gradientes se han originado como consecuencia de la actividad de una bomba iónica que consume energía.

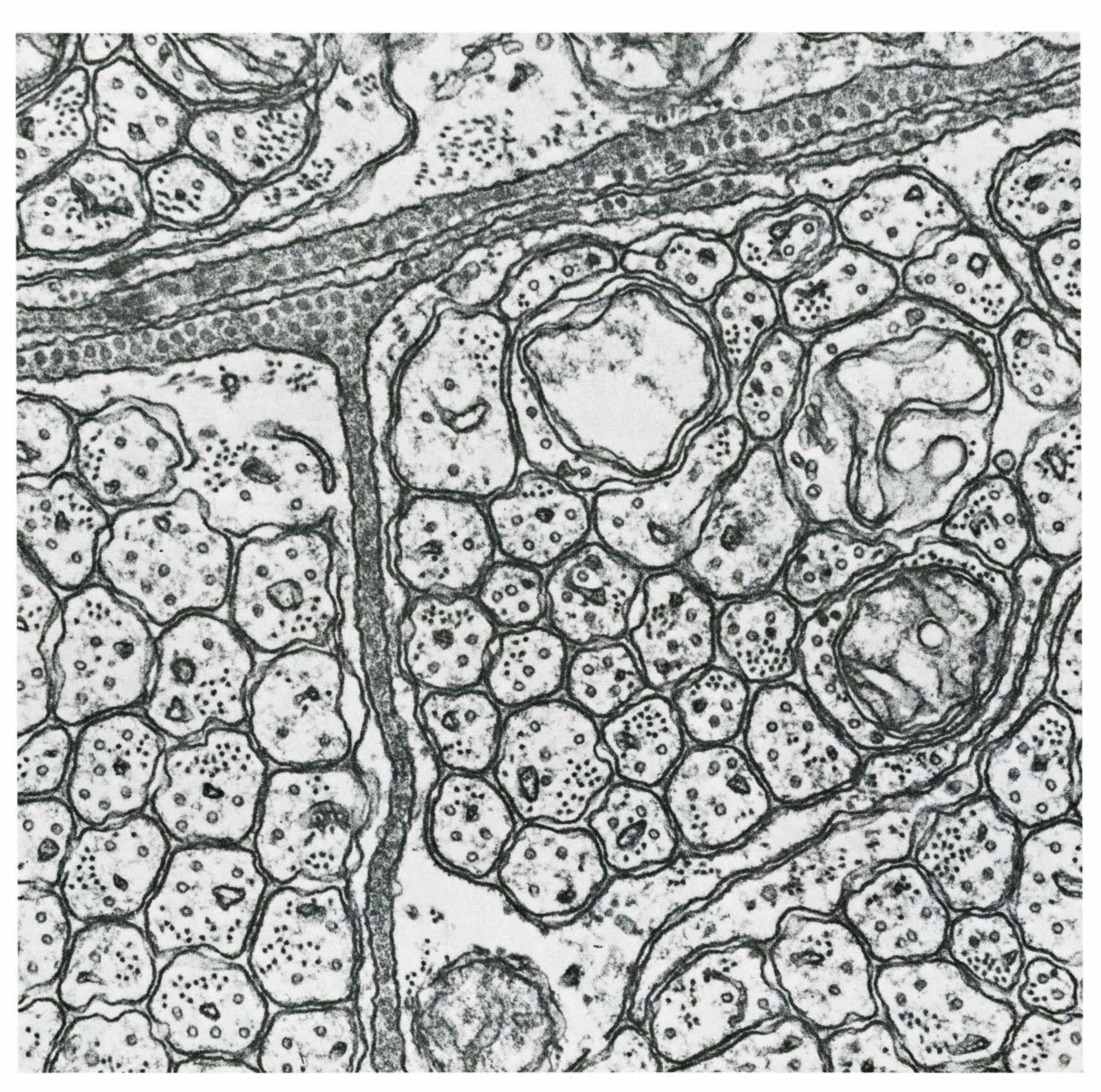
Como en muchas células vivas, el interior del axón –el axoplasma– es rico en iones potasio y pobre en iones sodio, mientras que en los líquidos corporales que rodean el axón las concentraciones relativas de estos iones están invertidas. Debido a esta desigual distribución de los iones y a la mayor permeabilidad de la membrana al potasio en condiciones de reposo, el nervio tiene un potencial de reposo de 60 a 70 milivolt, siendo el interior eléctricamente negativo con respecto al exterior.

El fenómeno que desencadena la actividad eléctrica del nervio es la despolarización de la membrana, es decir, la reducción de la diferencia de potencial a través de la misma, producida bien por una excitación procedente de otra célula nerviosa, bien por un impulso propagado desde un punto distante de la fibra. Esta despolarización abre canales en la membrana altamente selectivos para el sodio. Los iones sodio empiezan a fluir hacia el interior de la fibra, siguiendo su gradiente electroquímico. La entrada de iones con carga positiva provoca una despolarización adicional de la membrana y, en consecuencia, se abren más ampliamente las compuertas que controlan los canales de sodio. De esta forma, una vez se ha iniciado el incremento de permeabilidad al ion sodio, ésta aumenta de forma explosiva hasta que

todos los canales quedan abiertos. Como consecuencia de estos hechos el potencial de membrana primero cae a cero y luego se invierte. El flujo entrante de sodio disminuye cuando el potencial alcanza un valor próximo al del equilibrio para este ion, y cesa completamente debido al proceso llamado inactivación, que cierra inexorablemente los canales de sodio. Después de un corto período de latencia, se abre un segundo conjunto de canales con permeabilidad selectiva para el ion potasio, y este ion fluye al exterior siguiendo su gradiente electroquímico hasta que se restablece el potencial de reposo original.

Debido a la estructura de cable del axón, la onda de despolarización se extiende a lo largo del mismo a una velocidad de unos 20 metros por segundo. Para alcanzar esta velocidad, un axón de calamar tiene un diámetro de 600 micrometros, mientras que el axón mielinizado de rana tiene únicamente 12 micrometros. En un axón de calamar la elevación de la concentración interna de sodio durante un impulso único es muy pequeña, y sólo se pierde alrededor de una millonésima del potasio intracelular, no detectándose cambios en la velocidad de bombeo iónico consiguientemente. A largo plazo, sin embargo, el intercambio de iones que proporciona la fuente inmediata de energía para la propagación del potencial de acción conduce a modificaciones de la composición iónica que tienen que compensarse por la actividad de la bomba de sodio alimentada por la energía procedente del metabolismo celular.

Para explicar esta secuencia de fenó-



MANOJOS DE AXONES, o fibras nerviosas individuales del nervio olfatorio del lucio. La micrografía electrónica muestra una sección tranversa del nervio. Cada axón consiste en un tubo membranoso lleno de un fluido viscoso, el axoplasma. La membrana axónica posee canales moleculares independientes para el sodio y el potasio; el flujo de estos dos iones a través de la membrana origina el impulso nervioso. En el axón de lucio, los canales de sodio y potasio son poco abundantes, estando distribuidos con una densidad de sólo unos pocos canales por micrometro cuadrado de membrana. Esta micrografía electrónica amplía las fibras 80.000 diámetros. Fue realizada por Alexander von Muralt, de la Universidad de Berna, Suiza.

menos, Hodgkin y Huxley postularon la existencia de conjuntos paralelos de canales iónicos permeables selectivamente al sodio o al potasio y que serían controlados por el campo eléctrico existente a través de la membrana. Estos investigadores desarrollaron ulteriormente un método para hacer estudios cuantitativos de las propiedades de los canales de sodio y potasio. Aplicando la denominada técnica del "voltaje fijo" ("voltage clamp") provocaban modificaciones instantáneas del potencial de membrana manteniéndolo fijo en el nuevo nivel mientras determinaban el flujo de corriente iónica resultante.

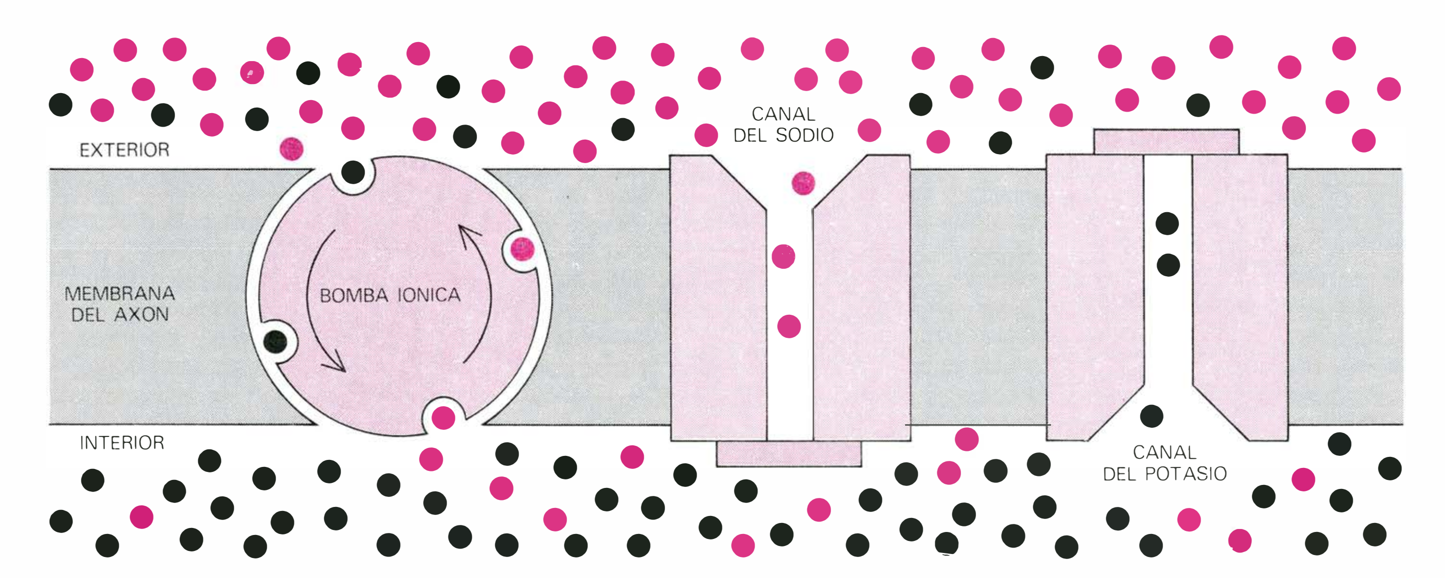
Desarrollando un concepto que idearon Kenneth S. Cole y George Marmont, su técnica del voltaje fijo suponía colocar dos electrodos en el axón, uno para detectar el potencial interno y el otro para permitir el paso de corriente a través de la membrana hacia un electrodo externo de grandes dimensiones. Se conectaba un amplificador de retroalimentación en el circuito, de forma que el potencial detectado podía compararse con el voltaje control fijado por el experimentador; y se hacía fluir la corriente necesaria en el axón para eliminar toda discrepancia entre el potencial interno y el potencial de control.

El potencial de membrana podía, pues, fijarse a un determinado nivel o secuencia de niveles, y era posible determinar la dependencia temporal y de voltaje de la corriente iónica en distintas condiciones experimentales. Desde que se introdujo la técnica del voltaje fijo ha sido adaptada para su aplicación a fibras nerviosas mielínicas y a muchos otros tipos de tejidos excitables, siendo el principal instrumento empleado en las investigaciones de las propiedades eléctricas de las membranas excitables.

Los datos obtenidos por Hodgkin y Huxley aplicando la técnica del voltaje fijo mostraron que los sistemas de sodio y potasio se comportaban de modo distinto cuando la membrana era despolarizada. La permeabilidad al sodio aumentaba rápidamente alcanzando un máximo y disminuía luego a cero cuando ocurría la inactivación, mientras que la permeabilidad al potasio aumentaba con una considerable latencia, pero no era inactivada cualquiera que fuera el tiempo de persistencia de la despolarización. El curso temporal de los cambios de permeabilidad fue analizado con detalle y los resultados expresados en un conjunto de famosas ecuaciones matemáticas que describían las propiedades de los dos tipos de canales iónicos en términos cuantitativos. Aplicando tales ecuaciones, Hodgkin y Huxley lograron su objetivo primario al comprobar que su análisis era acertado, pues podían predecir la forma precisa del potencial de acción y también otros parámetros tales como la velocidad de conducción y la cantidad de sodio y potasio intercambiados. Los cálculos fueron un completo éxito.

Estos trabajos que constituyen una pieza clásica en la investigación neurofisiológica representaron un gran paso adelante, pero dejaron sin respuesta dos cuestiones fundamentales. La primera cuestión planteaba el problema de cómo los canales iónicos pueden discriminar con extraordinaria finura entre iones monovalentes tan parecidos como el sodio, el potasio y otros metales alcalinos. Los iones de litio pasan por el canal del sodio casi con la misma facilidad que los iones sodio; sin embargo, los iones potasio atraviesan el canal con una dificultad doce veces superior, y la dificultad de paso de los iones rubidio y cesio es demasiado grande para ser medida. La segunda cuestión plantea el problema de la fuerte dependencia y adecuación temporal entre la apertura y cierre de los canales y el campo eléctrico a través de la membrana.

El mecanismo de la especificidad iónica de los canales ha sido estudiado por Bertil Hille, de la Universidad de Washington, en una elegante serie de experimentos en fibras mielínicas de nervios de rana utilizando la técnica del voltaje fijo. Hille estudió la permeabilidad del nodo de Ranvier a una serie de pequeños iones orgánicos positivos, demostrando que sólo aquellos que tenían una sección inferior a tres por cinco unidades angstrom eran capaces de atravesar los canales de sodio. La permeabilidad selectiva depende, pues, en parte del ajuste entre las dimensiones del ion penetrante y las del canal. No obstante, incluso entre iones positivos de tamaño adecuado había sorprendentes diferencias de permeabilidad. Por ejemplo, la hidroxilamina ($OH{-}NH_3^+$); y la hidracina ($NH_2{-}NH_3^+$) pasaban fácilmente a través del canal, mientras que la metilamina ($CH_3{-}NH_3^+$) no lo atravesaba.

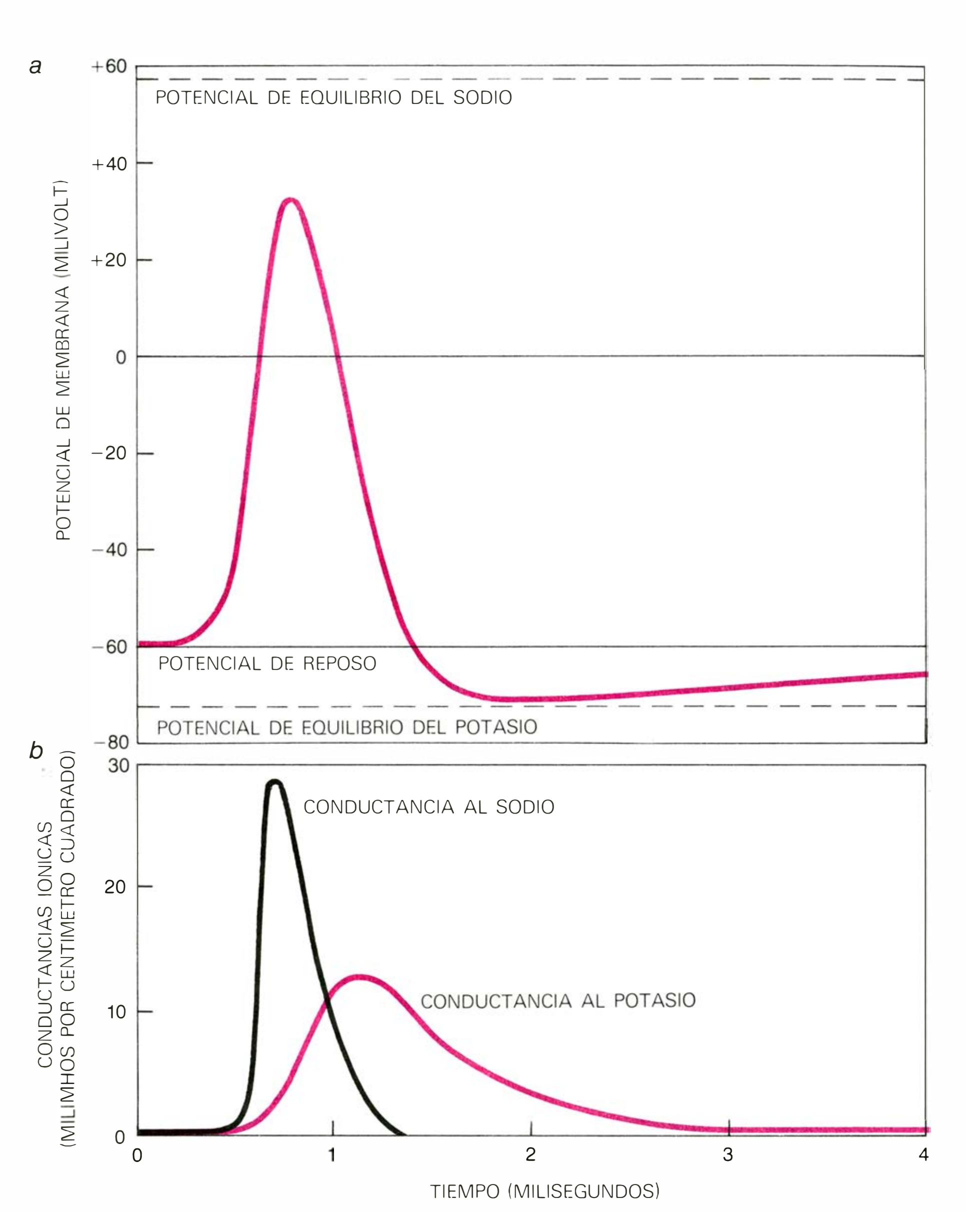


DOS TIPOS DE CANALES IONICOS atraviesan la membrana del axón. Una "bomba" impulsada por la hidrólisis del ATP transporta los iones sodio (*puntos de colores*) y los iones potasio (*puntos negros*) "cuesta arriba" y establece gradientes de concentración de estos iones a través de la membrana. El segundo tipo de canal permite el flujo de iones "cuesta abajo" en respuesta a cambios de voltaje de la membrana. En este diagrama, los canales independientes que corresponden al sodio y al potasio se muestran en su estado de reposo, con las compuertas iónicas cerradas debido al potencial de la membrana. Al reducir el potencial de reposo, los canales se abren (los de sodio, rápidamente, y los de potasio, lentamente) y originan un pulso de corriente que se propaga por el axón. Este artículo trata de los canales iónicos que permiten el movimiento de iones siguiendo su gradiente.

Para explicar esta discrepancia, Hille propuso que el canal del sodio en su punto más estrecho está rodeado por átomos de oxígeno, uno de los cuales posee carga negativa. Los iones positivos con grupos hidroxilo (OH) o amino (NH_2) pueden deslizarse a través del canal estableciendo enlaces de hidrógeno con el oxígeno, pero los grupos metilo (CH_3), que no forman enlaces de hidrógeno, son demasiado grandes para pasar por los canales.

Midiendo las corrientes iónicas cuando el sodio exterior es reemplazado por otros iones capaces o incapaces de atravesar el canal, Hille demostró la unión de los iones sodio a cargas negativas fijas de la pared del canal, por lo que estos iones lo atraviesan uno tras otro "en fila". Este comportamiento fue observado por Hodgkin y por el autor hace algunos años en experimentos, con iones marcados radiactivamente, practicados en los canales de potasio de los axones gigantes de sepia. Hille sugirió que existen cuatro barreras energéticas en el canal del sodio y que sólo un ion puede penetrar en el canal en un determinado momento.

La barrera energética mayor constituye el filtro de selección iónica. La relativa facilidad con que los iones hidratados de sodio sobrepasan la barrera depende de su habilidad para perder en etapas sucesivas las moléculas de agua estabilizantes al interaccionar con un grupo carboxílico ionizado (COO^-) de la pared del canal. La geometría del canal impide que los iones potasio algo mayores puedan desprenderse de su capa de hidratación con la misma facilidad, por lo que su barrera energética es mucho mayor. En la actualidad empieza a entreverse el detalle de la arquitectura molecular de la porción del canal iónico responsable de la selectividad.



EL IMPULSO NERVIOSO puede representarse mediante cambios de potencial a través de la membrana del axón (*a*) o mediante variaciones de la conductancia de la membrana a los iones sodio y potasio (*b*), ambos en una escala temporal de milisegundos. Durante el impulso, el interior de la membrana se vuelve transitoriamente positivo en relación al exterior. Luego se restablece el potencial de reposo.

Gran parte de los recientes progresos en la investigación de la relación existente entre el voltaje y el funcionamiento de los canales iónicos de sodio han sido posibles gracias a la ayuda de un veneno nervioso poco común llamado tetrodotoxina. Esta molécula se une a la apertura externa del canal del sodio bloqueando el paso de estos iones con tanta efectividad, que es una de las sustancias más tóxicas que se conocen. Se encuentra en grandes cantidades en los ovarios, hígado, piel e intestinos (y en menor proporción en el músculo) de los peces globo de la familia Tetraodontidae, de la cual la especie *Spheroides porphyreus* es conocida por los japoneses con el nombre de fugu. El fugu es una comida selecta, popular en el Japón, pero antes de comer el pescado es aconsejable cerciorarse de que el cocinero es un experto y conoce las partes realmente tóxicas que debe desechar. La tetrodotoxina está también presente en los tejidos de dos animales no relacionados con los peces globo: el tritón de California, *Taricha torosa,* y el pulpo australiano *Hapalochlaena maculosa.* Un análogo molecular que difiere sólo en un grupo lateral ha sido aislado de la rana de América Central *Atelopus chiriquensis.*

Otro veneno paralizante también de elevada toxicidad y que actúa de la misma forma que la tetrodotoxina, aunque posee una estructura algo distinta, es la saxitoxina. Esta molécula es elaborada por especies de plancton marino *Gonyaulax* que en condiciones favorables puede multiplicarse tan rápidamente que alcanza una densidad de muchos millones por litro, formando una "marea roja" que tiñe de color pardo rojizo la superficie del océano durante el día y le da luminiscencia por la noche. Los mariscos que se alimentan en aguas con marea roja concentran la toxina en sus tejidos y resultan venenosos para el hombre; una parálisis fatal puede resultar de la ingestión de cantidades tan pequeñas como un miligramo de saxitoxina. La formación independiente de moléculas tan poco comunes como la tetrodotoxina y la saxitoxina en varias ocasiones durante el proceso de la selección natural proporciona un ejemplo sorprendente del resultado de la evolución convergente.

La tetrodotoxina es una compleja molécula orgánica constituida por varios anillos enlazados y que posee en un ex-

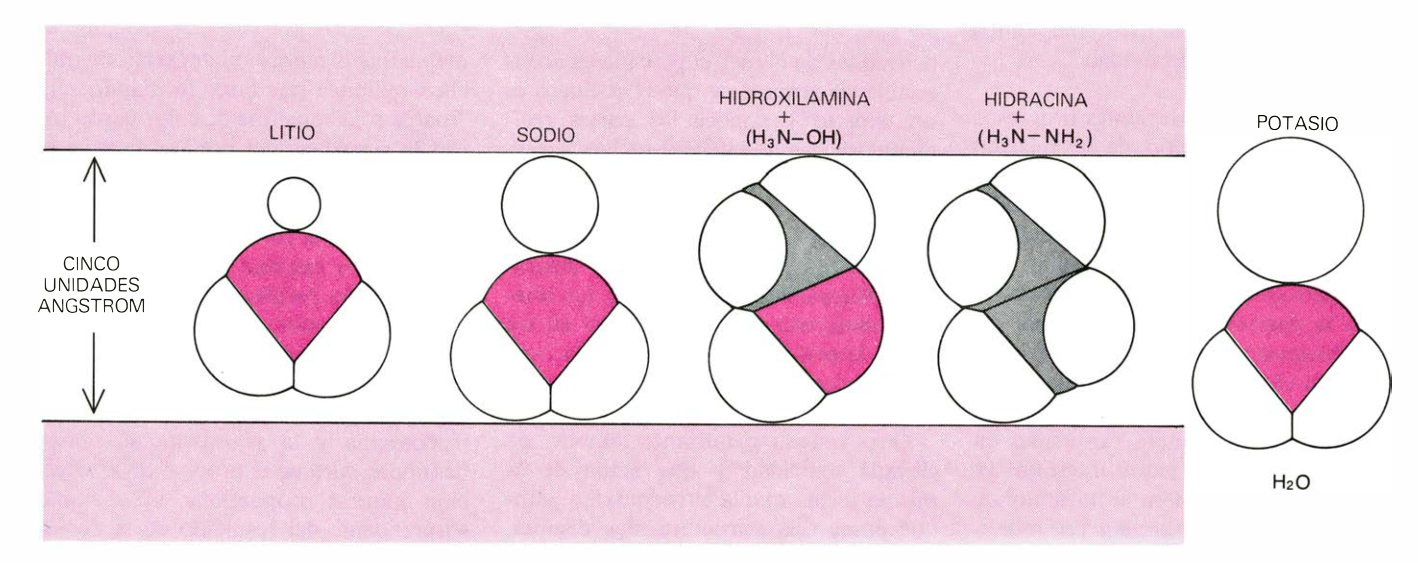
tremo un grupo guanidinio. Parece probable que el grupo guanidinio mimetice al ion sodio con carga similar y permita a la molécula entrar en la boca externa del canal del sodio donde se atasca, como un hombre obeso atrapado en medio de la portilla de un buque, taponando completamente el canal. Para explicar el porqué la tetrodotoxina actúa a una concentración tan pequeña hay que suponer que el resto de la molécula se adapta con precisión a la superficie externa del canal. El hecho de que la saxitoxina se una con la misma afinidad poseyendo esta molécula dos grupos guanidinios y diferente número de anillos sugiere que la molécula ocupa un sitio vecino, pero no un receptor idéntico. El receptor de la tetrodotoxina está situado exclusivamente en la parte externa del canal, puesto que ni siquiera una disolución muy concentrada de la misma tiene efecto bloqueante al ser introducida en el interior del axón. La tetrodotoxina y la saxitoxina son completamente específicas para el canal del sodio; ninguna de ellas incide lo más mínimo en la conductancia del canal del potasio.

Cuando un axón es expuesto a la tetrodotoxina o a la saxitoxina, sólo una molécula de toxina se une a un canal iónico, por lo que estas moléculas pueden ser utilizadas para determinar el número de canales de la membrana. Esta posibilidad fue utilizada en primer lugar por John Moore, Toshio Narahashi y Trevor Shaw, de la Universidad de Duke, quienes establecieron un ingenioso bioensayo para la determinación indirecta del número de sitios de unión de la tetrodotoxina a los pequeños axones del nervio de la pata de langosta. Estos autores demostraron que los canales de sodio en este tipo de fibra nerviosa son escasos y están muy alejados unos de otros, habiendo sólo unos trece por cada micrometro cuadrado de membrana.

Esta observación fue ulteriormente confirmada y generalizada a otros tipos de axón por J. Murdoch Ritchie y sus colaboradores en la Facultad de Medicina de la Universidad de Yale. Empleando tetrodotoxina y saxitoxina marcadas con tritio, el isótopo radiactivo del hidrógeno, fue posible determinar directamente la cantidad de toxina unida. Los resultados de estos autores demostraron que los axones más pequeños, como los de los nervios olfatorios de los peces, tienen el menor número de canales de sodio por micrometro cuadrado de membrana, sólo alrededor de tres, mientras que las fibras algo mayores del nervio vago del conejo tienen unos 25 canales por micrometro cuadrado. En la Asociación de Biología Marina de Plymouth, Hans Meves y Rock Levinson demostraron que el axón gigante de calamar tenía una densidad de canales de sodio mucho mayor: hasta unos 500 por micrometro cuadrado. Esta cifra elevada coincidía con otros datos, especialmente con los que cabía esperar de la magnitud de la corriente de la compuerta iónica del sodio. El número de canales no es muy diferente del calculado por Hodgkin en relación con la velocidad máxima de conducción del impulso nervioso en un axón de 500 micrometros.

Un segundo aspecto en el que la tetrodotoxina ha demostrado ser un instrumento de valor extraordinario para los electrofisiólogos es el de contribuir al esclarecimiento del mecanismo de control de las compuertas de los canales iónicos. Tal como Hodgkin y Huxley pusieron de manifiesto en 1952, la fuerte dependencia entre la permeabilidad iónica de la membrana y la diferencia de potencial a través de la misma indica que los canales de sodio y potasio deben estar controlados por movimientos de cierre y apertura de las "compuertas" de los canales, que se comportan como si poseyeran una carga de considerable magnitud. En condiciones de reposo, con un potencial de unos −65 milivolt, la mayoría de las partículas con carga de la compuerta iónica del sodio están en su forma de reposo o "configuración de cierre", y el canal permanece cerrado. En respuesta a un desplazamiento del potencial de membrana hacia un valor más positivo, es decir, hacia la despolarización, las partículas de la compuerta iónica experimentan un cambio adoptando la "configuración abierta" y permitiendo que los iones atraviesen el canal.

El funcionamiento del sistema de compuertas de los canales iónicos supone un movimiento de cargas en un campo eléctrico y, por ello, el cambio de configuración descrito previo a la apertura del canal originará un pequeño desplazamiento de carga a través de la membrana. En otras palabras, tendría que detectarse un flujo fugaz de corriente debido al desplazamiento de cargas precediendo el paso de la corriente iónica. Hodgkin y Huxley predijeron la existencia de esta corriente de compuerta, pero no pudieron medirla por la



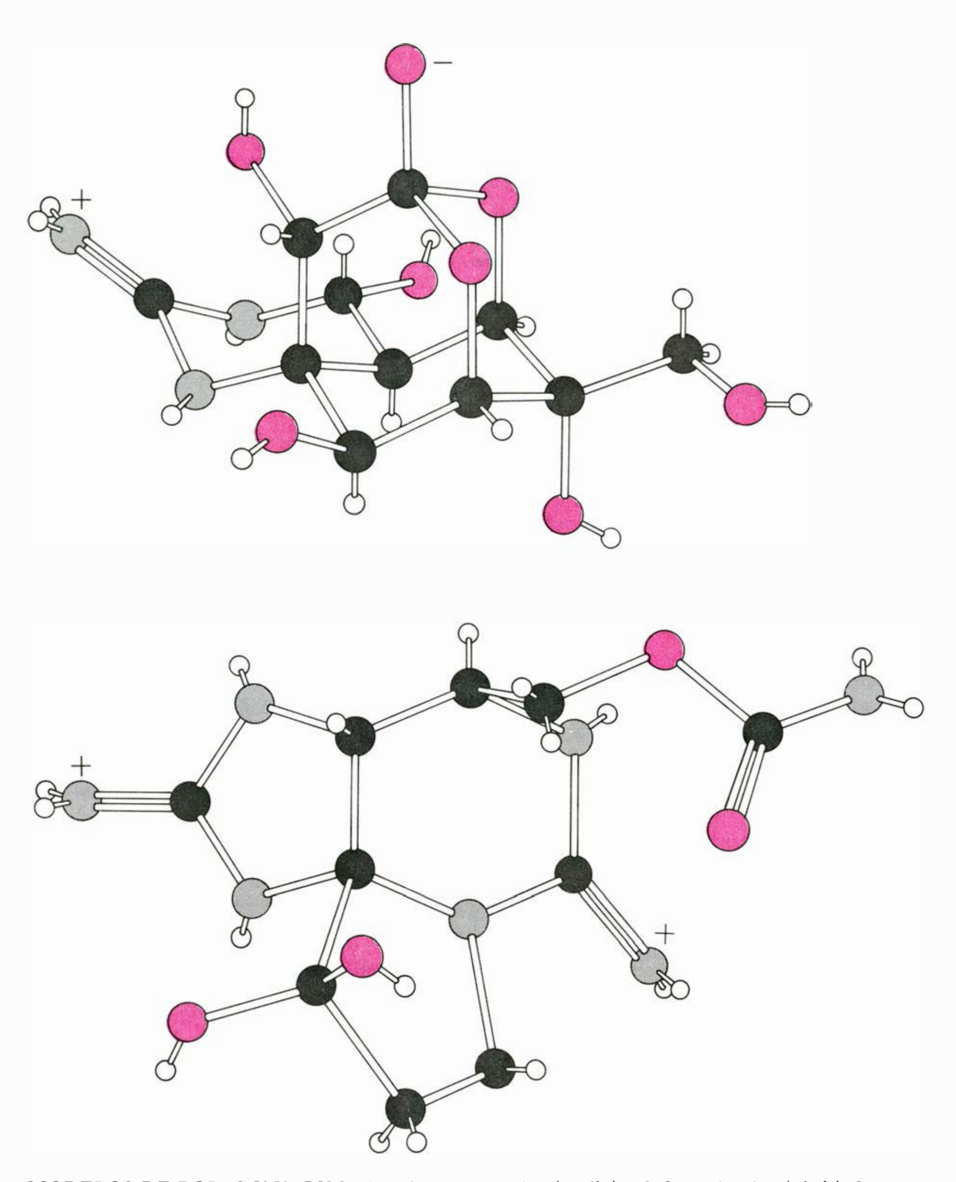
SELECTIVIDAD IONICA del canal del sodio. Constituye una consecuencia de sus dimensiones y de sus propiedades electrostáticas. Los iones litio parcialmente hidratados, o los iones sodio, así como la hidroxilamina y la hidracina, pequeñas moléculas cargadas positivamente, tienen las dimensiones apropiadas para atravesar el canal del sodio, que parece tener una sección transversal de tres por cinco unidades angstrom en su punto más estrecho. Un ion potasio hidratado, sin embargo, es demasiado grande para atravesar el canal. Las moléculas se representan con modelos de tipo espacial.

simple razón de que en aquel tiempo no había la posibilidad de bloquear la entrada de iones a través del canal de sodio, que enmascaraba completamente la citada corriente. En el corto período de tiempo en el que un canal de sodio permanece abierto durante el paso de un impulso nervioso, alrededor de 100 iones sodio pasan por el canal correspondiente; durante un largo pulso de voltaje fijo, hasta 500 iones pueden atravesar el canal. La apertura de la compuerta del canal, por otra parte, supone la transferencia de hasta un máximo de cuatro cargas electrónicas en el interior de la membrana.

Puesto que la corriente de compuerta es mucho más pequeña que la corriente iónica, no puede detectarse fácilmente si no se bloquean completamente los canales iónicos. Disponiendo de la tetrodotoxina, las condiciones resultan más favorables para el registro de la corriente de compuerta iónica gracias a la afortunada circunstancia de que cuando la toxina está unida a la boca del canal, bloqueándolo, la apertura y cierre de la compuerta del canal no se ve afectada y la corriente de compuerta sigue fluyendo.

Si la corriente de sodio en un axón de calamar es abolida por la aplicación de una gran dosis de tetrodotoxina (y también reemplazando el ion sodio en el medio extracelular con tampón "Tris"), y al mismo tiempo los canales del potasio son bloqueados desde el interior perfundiendo el axón con iones cesio o tetraetilamonio, la corriente de compuerta constituye una fracción sustancial del desplazamiento de corriente registrado cuando el potencial fijado en la membrana es modificado bruscamente. Utilizando este enfoque, las corrientes de compuerta de los canales del sodio fueron observadas por primera vez durante la estación del calamar en 1972-73 por Clay Armstrong y Pancho Bezanilla en el laboratorio de Biología Marina de Woods Hole, Mass., y, poco después, por Eduardo Rojas y el autor en Plymouth. Como en otros casos, la nueva observación en el axón gigante de calamar fue seguida de la demostración de su existencia en otros tipos de nervios, llevándose a cabo con éxito estudios de las corrientes de compuerta en el nodo de Ranvier del nervio de rana por Wolfgang Nonner, Berthold Neumcke y Robert Stämpfli, de la Universidad de Homburg, y en otros laboratorios.

Aunque con la ayuda de la tetrodotoxina la corriente de compuerta puede separarse de la corriente iónica sin dificultad, el problema estriba en distinguir

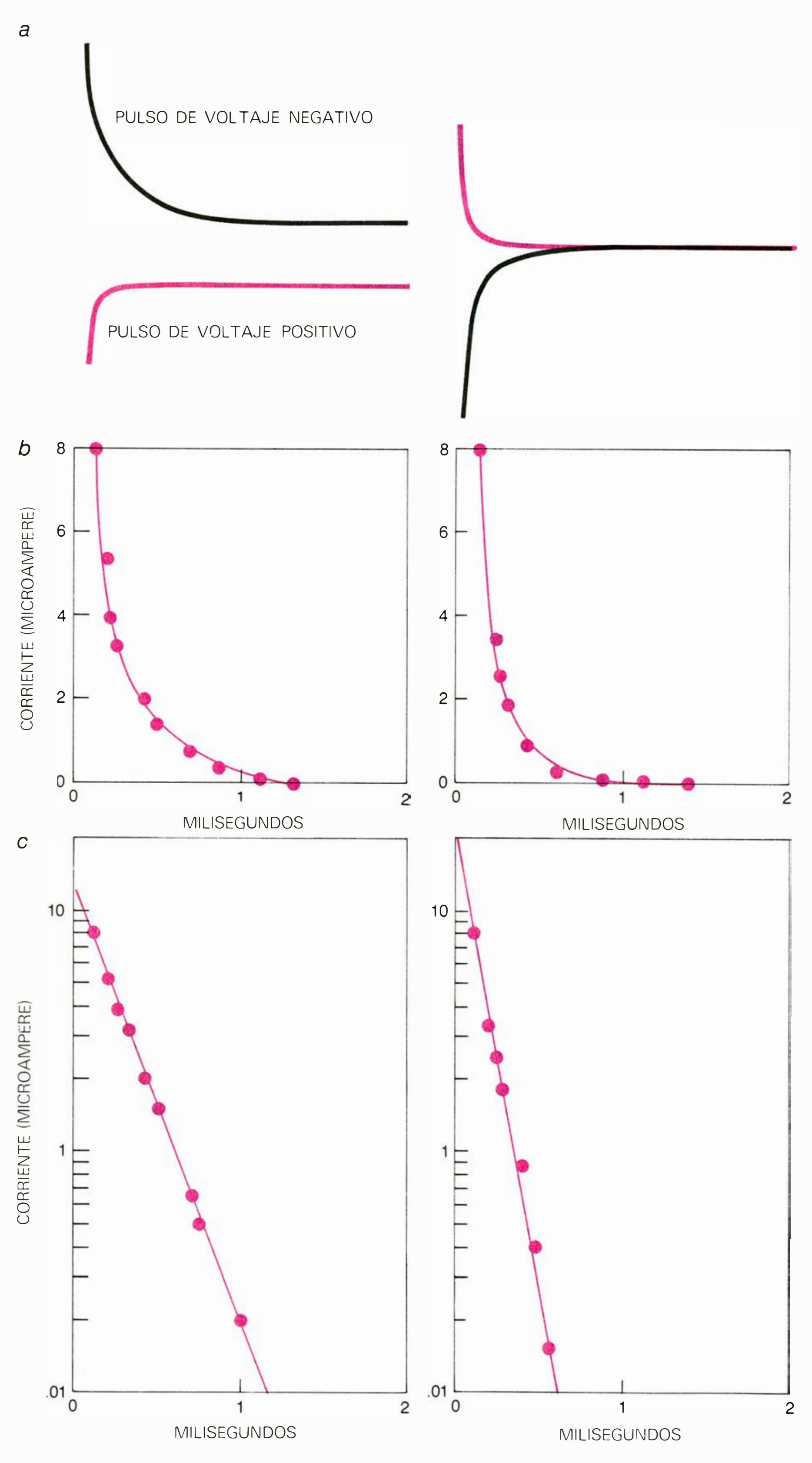


MODELOS DE BOLAS Y VARILLAS de la tetrodotoxina (*arriba*) y de la saxitoxina (*abajo*). Se trata de dos potentes venenos nerviosos que se unen fuerte y específicamente a los canales del sodio. Los modelos revelan algunas características estructurales comunes. En ambas moléculas un grupo cargado positivamente mimetiza la carga positiva del sodio entrando en el canal, pero la voluminosa estructura formada por múltiples anillos impide que la toxina lo atraviese. Las dos toxinas hicieron posible el estudio.

entre la corriente de desplazamiento originada por la carga y descarga de la gran capacidad estática de la membrana y la corriente de desplazamiento producida por los movimientos de las partículas de las compuertas. Puede asumirse razonablemente que para la oscilación de la capacidad, la cantidad total de carga transferida a través de la membrana al aplicar un pulso de voltaje fijo varía en proporción directa a la magnitud del cambio de potencial. Por tanto, las oscilaciones de capacidad para pulsos de igual amplitud y sentido contrario serán siempre exactamente simétricas.

Las partículas involucradas en el mecanismo de compuerta de los canales iónicos estarán inicialmente distribuidas de forma desigual entre sus dos configuraciones, y para un número total constante, la transferencia de carga cuando cambian de estado mostrará saturación si los pulsos son de suficiente magnitud. La amplitud de la corriente de compuerta será, por tanto, asimétrica con respecto al potencial. Partiendo de un potencial negativo constante en el cual las partículas están en la "configuración cerrada", el registro de la corriente para un pulso de voltaje fijo de hiperpolarización (sentido negativo) consistirá en un cambio puro (oscilación) de capacidad, puesto que las compuertas iónicas no pueden estar más cerradas de lo que están. La corriente de desplazamiento para un pulso de despolarización (sentido positivo) de la misma amplitud consistirá, sin embargo, en una oscilación de capacidad (igual y de signo opuesto) más la corriente de apertura generada al abrirse las compuertas iónicas de los canales.

El método de registro de las corrien-

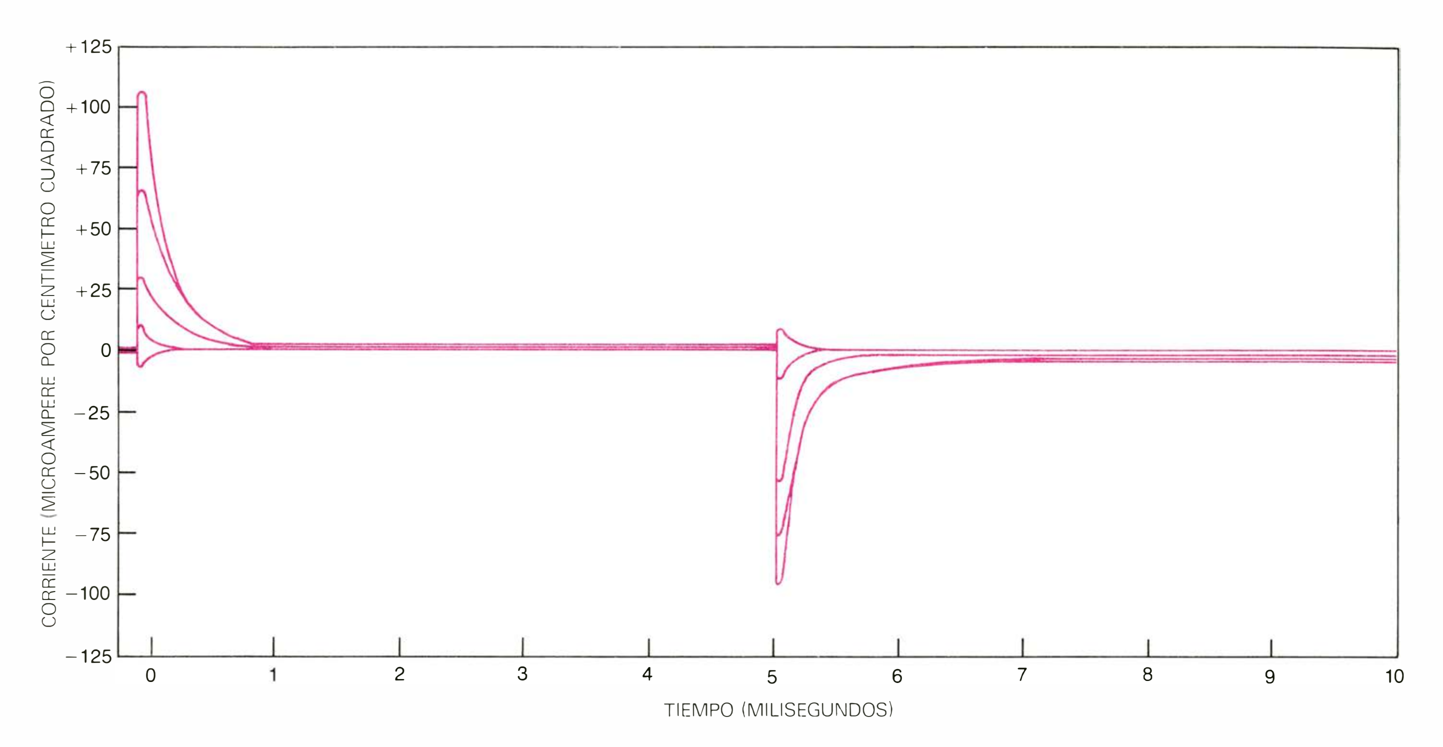


CORRIENTE DE COMPUERTA, asociada al movimiento de partículas o subunidades con carga dentro del canal del sodio. Las subunidades abren o cierran el canal en respuesta a variaciones del potencial de la membrana. La corriente de compuerta puede detectarse únicamente después de bloquear la corriente del sodio con tetrodotoxina o la del potasio con cesio y después de neutralizar la oscilación momentánea simétrica de la capacidad de la membrana. En estas condiciones, la aplicación a través de la membrana de pulsos de voltaje iguales y opuestos da lugar a flujos de corriente ligeramente distintos; la discrepancia constituye la corriente de compuerta. Los trazos en *a* son registros de barrido simple de la corriente de la membrana para pulsos de voltaje de ± 120 milivolt. La diferencia entre ellos está representada en la parte inferior en forma lineal (*b*) o logarítmica (*c*). Un desplazamiento del potencial en sentido positivo (pero no en sentido negativo) determina una breve corriente de compuerta hacia fuera que coincide con la apertura de los canales de sodio. La corriente de compuerta cae exponencialmente a cero cuando las subunidades de compuerta terminan por alcanzar su nueva configuración.

tes de compuerta inicialmente utilizado por Armstrong y Bezanilla en Woods Hole y por Meves y nosotros en Plymouth consistió simplemente en sumar, con un aparato que calculaba la media de las señales, los registros de corriente para pulsos de igual voltaje y de sentido opuesto en axones con canales iónicos bloqueados, anulando automáticamente la oscilación de la capacidad. Esta técnica ha sido después mejorada en ciertos aspectos empleando una computadora en lugar del aparato utilizado anteriormente para promediar las señales. Pero el principio sigue siendo el mismo. La corriente de compuerta aumenta formando un pico inmediatamente después de la aplicación de la variación de potencial, y, a partir de ahí, cae exponencialmente hasta cero a medida que las partículas de la compuerta del canal alcanzan su nueva configuración. Como muestra la ilustración superior de la página 79, al aumentar la amplitud del pulso se transfiere más carga, pero el tiempo de relajación disminuye.

Partiendo de las propiedades del estado de equilibrio de las partículas implicadas en la apertura del canal, tal como sugirieron en principio Hodgkin y Huxley en base a la prueba cinética, el mecanismo de compuerta implica la cooperación de tres (o posiblemente cuatro) partículas cargadas en cada canal. Cada partícula se comporta como si acarreara una sola carga electrónica, pero podría tratarse de un cierto número de grupos laterales con carga desplazándose a través de una pequeña fracción del campo eléctrico. A partir de los resultados experimentales, no resulta posible determinar la disposición de los grupos laterales dentro de la partícula. Comparando la velocidad con que las partículas alcanzan su nueva configuración y la velocidad de paso de los iones a través de los canales, resulta evidente que la cinética de la corriente ligada a las compuertas iónicas no puede conciliarse directamente con las ecuaciones de Hodgkin y Huxley. A fin de que los datos encajen, hay que asumir la existencia de una "configuración activada" entre las configuraciones "cerrada" y "abierta", y que, una vez las partículas han alcanzado la configuración activada, tendría lugar un período de interacción cooperativa de las partículas antes de que los canales se abran. Durante este período no fluye corriente adicional.

La otra conclusión importante se refiere al mecanismo de inactivación de los canales de sodio, es decir, al proceso que cierra de nuevo los canales aproxi-

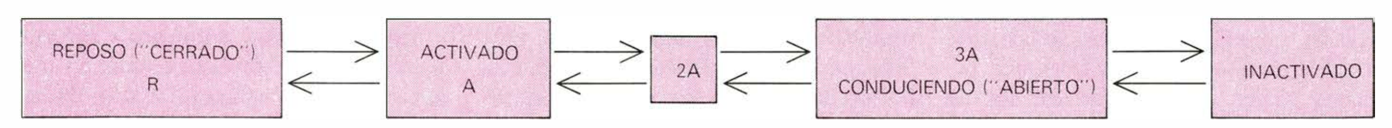


CONJUNTO DE CORRIENTES DE COMPUERTA registrado en un axón gigante de calamar por el autor con George Malachowski y Dirk van Helden en los laboratorios de la Asociación de Biología Marina de Gran Bretaña en Plymouth. El potencial de membrana se mantuvo en −100 milivolt y se aplicaron pulsos de −120, −80, −40, 0 y +40 milivolt. Al desplazar el voltaje en sentido positivo, la corriente de compuerta fluía hacia fuera.

madamente un milisegundo después de su apertura, aunque la membrana permanezca despolarizada. Unánimemente se desecha la idea de que, además de las partículas con carga que determinan la apertura de los canales, exista otro grupo de partículas encargadas de su bloqueo; estas partículas cerrarían el canal después de un período de latencia adecuado. La razón de esta conclusión estriba en que las partículas determinantes del bloqueo deberían estar también cargadas y no es posible registrar ninguna corriente de desplazamiento asimétrico que correspondiera a su movimiento. Armstrong y Bezanilla han propuesto que la inactivación implica una reestructuración de las partículas de las compuertas iónicas que lentamente conduce al cierre de los canales abiertos. Se cree que el cambio de configuración final va acompañado de un movimiento muy pequeño de carga o de ningún movimiento, siendo, pues, eléctricamente mudo. Aparentemente deja las partículas de la compuerta en un estado de inmovilización parcial, determinando que con la repolarización de la membrana su vuelta al estado de reposo sea demasiado lento como para que se registrara con facilidad.

El estudio de la corriente ligada al movimiento de las compuertas de los canales del ion sodio ha conducido al establecimiento del concepto de la existencia de una parte del canal iónico sensible al voltaje, que estaría constituida por un sistema de múltiples subunidades que pasan por múltiples estados de configuración. Puede elaborarse un modelo matemático que incorpore los principales hallazgos descritos, pero probablemente resultaría poco atrayente para los investigadores, puesto que todos los modelos con multiplicidad de estados implican demasiados parámetros arbitrarios para proporcionar una solución única del problema. No obstante, no puede negarse que funcionalmente los canales de sodio pueden encontrarse (1) en reposo, (2) conduciendo o (3) inactivados o refractarios, de modo que, guste o no guste, el sistema debe existir, al menos, en tres estados distintos. Los estados intermedios adicionales pueden resultar también necesarios para explicar la cronología de las transiciones entre los estados principales. Es de esperar que, a partir de determinaciones futuras más precisas de la corriente de compuerta, se disponga de una mayor información sobre dichas transiciones y puedan ser caracterizadas independientemente de los datos derivados de la cinética global del sistema. La ausencia de tales datos debe tenerse en cuenta a la hora de elaborar modelos, si se pretende que éstos sean realistas.

La elaboración de un modelo molecular plausible del canal del sodio aún está más lejos de nuestro alcance en el momento presente. Parece claro que el filtro determinante de la permeabilidad selectiva del canal está localizado cerca



MODELO DE LOS ESTADOS MULTIPLES del canal del sodio propuesto por el autor tomando como base los recientes estudios sobre las corrientes de compuerta de los canales iónicos. El modelo muestra una secuencia hipotética de estados configuracionales a través de los cuales pasan las subunidades que constituyen la compuerta del canal durante los movimientos de apertura y cierre. En respuesta a una desviación positiva del voltaje a través de la membrana del axón, las subunidades de la compuerta iónica experimentan un cambio de configuración hacia el estado activado. A continuación interaccionan secuencialmente; cuando las tres subunidades que controlan un determinado canal han interaccionado, el canal se abre y permite el paso de los iones sodio. Después de permanecer abierto un breve período de tiempo, el canal se cierra al experimentar un cambio de configuración que lo conduce al estado inactivado. La doble flecha entre los estados indica que cada reacción puede tener lugar en ambos sentidos. En la práctica, la mayoría de los canales pasan a través del ciclo completo desde el estado de reposo al estado inactivado y, nuevamente, al estado de reposo con cada impulso. Una corriente de compuerta mensurable acompaña sólo el paso del estado de reposo al activado o el inverso.

de su boca externa, y probablemente el punto más estrecho del canal es una estructura rígida que no se modifica por cambios de potencial de la membrana. La dificultad estriba en describir la parte central del canal donde está localizada la compuerta, puesto que no hay precedentes que permitan ni siquiera imaginar cómo una alteración del campo eléctrico a través de una macromolécula puede abrir una vía acuosa a su través, y mucho menos un mecanismo cuya apertura y cierre muestra una cinética tan compleja. Cualquier descripción del funcionamiento interno del canal del sodio es, pues, pura fantasía.

Los comentarios anteriores se han referido casi exclusivamente a los canales de sodio, lo cual no significa que los canales de potasio sean menos importantes. Existen dos razones por las cuales estos canales no son bien conocidos. En primer lugar, no se dispone de ningún agente bloqueante de los canales de potasio que posea una especificidad parecida a la de la tetrodotoxina para los canales de sodio. Si bien los iones de tetraetilamonio cargados positivamente pueden entrar en los canales de potasio desde el lado interno y taponarlos de forma efectiva, se requiere una concentración elevada para su completo bloqueo. Por tanto, el truco de aplicar un marcador radiactivo a las moléculas bloqueantes para determinar el número de canales de potasio no puede aplicarse en este caso. En segundo lugar, debido a que los canales de potasio se abren con retraso y con una lentitud diez veces mayor que la de los canales de sodio, la corriente de compuerta del potasio es mucho más difícil de detectar.

En un aspecto, sin embargo, la lentitud de los canales de potasio puede representar una ventaja, puesto que da como resultado su mayor contribución al "ruido" eléctrico de fondo originado por las fluctuaciones al azar de las corrientes iónicas que fluyen a través de la membrana. El examen del espectro del ruido de fondo en diferentes tipos de membranas excitables constituye un campo de estudio totalmente nuevo, que promete proporcionar importante información sobre algunas de las características de los canales iónicos a las que no resulta posible aplicar otras técnicas de investigación. Los estudios actuales llevados a cabo por Charles F. Stevens, de la Facultad de Medicina de la Universidad de Yale, y por Franco Conti, del Consejo Nacional de Investigación de Italia, entre otros, utilizando la citada técnica, han proporcionado las mejores estimaciones de que disponemos actualmente sobre la conductancia de un canal iónico y sobre el número de canales de potasio en el axón gigante de calamar. Estos resultados indican que hay sólo unos 50 canales de potasio por micrometro cuadrado de membrana axónica, aproximadamente un canal de potasio por cada 10 canales de sodio. La escasez relativa de canales de potasio reduce aún más la magnitud probable de la corriente de compuerta del canal iónico del potasio.

Demasiados problemas para ser resueltos contando únicamente con los resultados de las determinaciones electrofisiológicas. Puesto que los cambios dependientes del voltaje en las conductancias iónicas de la membrana de la célula nerviosa están determinados por reorganizaciones estructurales dentro de los canales iónicos, surge la cuestión de si dichos cambios pueden detectarse utilizando otras técnicas aparte de las eléctricas. Con esta idea, algunos investigadores, incluido el autor, han buscado cambios pasajeros en varios parámetros que pueden estar asociados con la configuración molecular de las partículas de las compuertas. Tales parámetros comprenden propiedades ópticas de la membrana, cantidad de luz dispersada por el axón, emisión y absorción de luz después de teñir la membrana con un colorante fluorescente y la liberación y absorción de calor por el axón. Aunque en cada caso los efectos pueden observarse durante el paso de impulsos nerviosos o la aplicación de pulsos de voltaje fijo a la membrana, se llegó desafortunadamente a la conclusión de que tales cambios reflejaban siempre la respuesta de la membrana como un todo, y no estaban específicamente relacionados con fenómenos que tuvieran lugar dentro de los canales iónicos.

Para llegar a una comprensión completa de los mecanismos de funcionamiento de los canales iónicos es necesario estudiar su bioquímica. Puede asumirse razonablemente que, al igual que otras macromoléculas implicadas en el transporte de iones a través de las membranas celulares, los canales iónicos son proteínas de gran tamaño. El canal proteico posee probablemente una porción hidrofóbica central, que se encuentra inmersa en el ambiente lipídico que le ofrecen las colas hidrocarbonadas de los fosfolípidos de la membrana y porciones hidrofílicas proyectadas en el medio acuoso a ambos lados de la membrana. Como en otras proteínas de la membrana, esta doble naturaleza hace difícil aislar y purificar tales macromoléculas para ser sometidas a un análisis estructural como la difracción de rayos X, puesto que el mismo proceso de extracción posiblemente las altera e inactiva. Además, para poder aplicar técnicas de purificación tales como la cromatografía de afinidad, se requiere una molécula adecuada a la que la proteína del canal se una selectivamente. La tetrodotoxina sería la sustancia de elección, pero hasta que recientemente Michel Lazdunski, de la Universidad de Niza, logró preparar con éxito un derivado fluorescente de la tetrodotoxina, la sustancia resultaba muy difícil de manejar químicamente.

La clave del éxito en este tipo de trabajo estriba a menudo en encontrar un material de partida con una gran riqueza de la proteína deseada, tal como ejemplifica la explotación por Jean Pierre Changeaux, del Instituto Pasteur de París, del órgano eléctrico de la raya *Torpedo* como fuente de la proteína receptora de la acetilcolina. En nuestro laboratorio otro tipo de órgano eléctrico ha merecido especial atención, el de la anguila *Electrophorus.* Aunque las electroplacas de *Electrophorus* no sean tan ricas en canales de sodio como las de *Torpedo,* lo son en receptores de acetilcolina; una anguila grande proporciona un kilogramo o más de tejido eléctrico. A partir de investigaciones sobre la pérdida de capacidad de unión de tetrodotoxina en preparaciones microsómicas de canales de sodio que han sido bombardeadas con electrones de alto voltaje, mis colegas Clive Ellory y Levinson determinaron el peso molecular de la proteína del canal, obteniendo un valor de 230.000 dalton. Trabajando con Michael A. Raftery en el Instituto de Tecnología de California, Levinson y otros han podido extraer del órgano eléctrico de *Electrophorus* una proteína casi pura que se une a la tetrodotoxina y que tiene un peso molecular del mismo orden al previamente calculado. Aunque aún quedan obstáculos técnicos de gran envergadura por vencer, los electrofisiólogos confían poder estudiar los canales del sodio incorporados en membranas artificiales.

En la actualidad, comienza a emerger un tenue esbozo de la organización molecular de los canales iónicos de las membranas excitables. Investigaciones posteriores permitirán determinar de forma precisa la estructura y el funcionamiento de los canales iónicos, lo que sin duda constituirá uno de los avances más importantes para el esclarecimiento de los mecanismos de la excitabilidad biológica.

Delfines

Estos descendientes de mamíferos terrestres que se adaptaron al mar tienen un cerebro grande, aprenden rápidamente y exhiben un rico repertorio vocal. Pero la falta de pruebas deja en suspenso la cuestión de su supuesta inteligencia

Bernd Würsig

Debido a que tienen un cerebro grande, aprenden rápidamente los trucos ideados por sus entrenadores humanos y exhiben un rico repertorio de señales vocales, se cree generalmente que los delfines tienen un nivel de inteligencia no superado por ningún otro animal y que quizás es incluso similar al del hombre. En base a las observaciones de delfines que he realizado en varios de sus hábitats naturales, creo que el esfuerzo que existe para situarlos firmemente en el espectro de inteligencia animal es prematuro. En el estado actual de la tecnología marina es absolutamente imposible que los observadores humanos pasen más que momentos breves y aislados con unos animales que viven en el océano y se desplazan rápidamente a lo largo de grandes distancias. Cuando se conozca mucho más sobre el comportamiento de los delfines, la cuestión acerca de su inteligencia se responderá por sí misma. En la actualidad, lo más que puede decirse es que los delfines son animales gregarios que forman manadas, comparables por su comportamiento individual y social a los mamíferos terrestres, más fácilmente observables, que forman rebaños y manadas.

Los delfines evolucionaron hace, por lo menos, 50 millones de años a partir de mamíferos terrestres que pudieron haberse parecido a los ungulados artiodáctilos de hoy día, como el ganado vacuno, los cerdos y los búfalos. Después de penetrar en el mar, los delfines se adaptaron cada vez mejor a la vida en el agua: su pelaje ancestral fue sustituido por una gruesa capa de grasa gelatinosa, se hicieron lisos e hidrodinámicos, perdieron sus extremidades posteriores, con excepción de algunos restos internos, y desarrollaron una poderosa cola, al tiempo que sus extremidades anteriores se modificaban en aletas nadadoras y, al parecer como una ulterior ayuda para la navegación y la estabilización, muchas especies desarrollaron una aleta dorsal.

Pueden identificarse más de 30 especies de delfines. Pertenecen al suborden Odontocetos (cetáceos con dientes) del orden Cetáceos. (Las marsopas, relativamente parecidas y que a veces se confunden con los delfines, son también miembros del suborden; se distinguen de los delfines sobre todo por tener un hocico menos puntiagudo y también por sus dientes comprimidos lateralmente, espatulados. Debe señalarse, sin embargo, que muchos estudiosos norteamericanos de los mamíferos marinos llaman marsopas a todos los cetáceos odontocetos pequeños, sin tener en cuenta sus características físicas.) Unas cuantas especies de delfines viven en las aguas dulces, pero la mayoría de especies moran en el océano. Las especies dulceacuícolas se desplazan en pequeños grupos o son casi solitarias, mientras que las especies oceánicas (como el delfín moteado del Pacífico) pueden congregarse en bancos de varios miles de individuos. Estas cifras recuerdan las manadas de búfalos de Norteamérica y las de animales herbívoros de las llanuras del Serengeti, en Africa; ello hace pensar en las semejanzas de comportamiento y ecológicas entre los delfines y sus lejanos parientes terrestres.

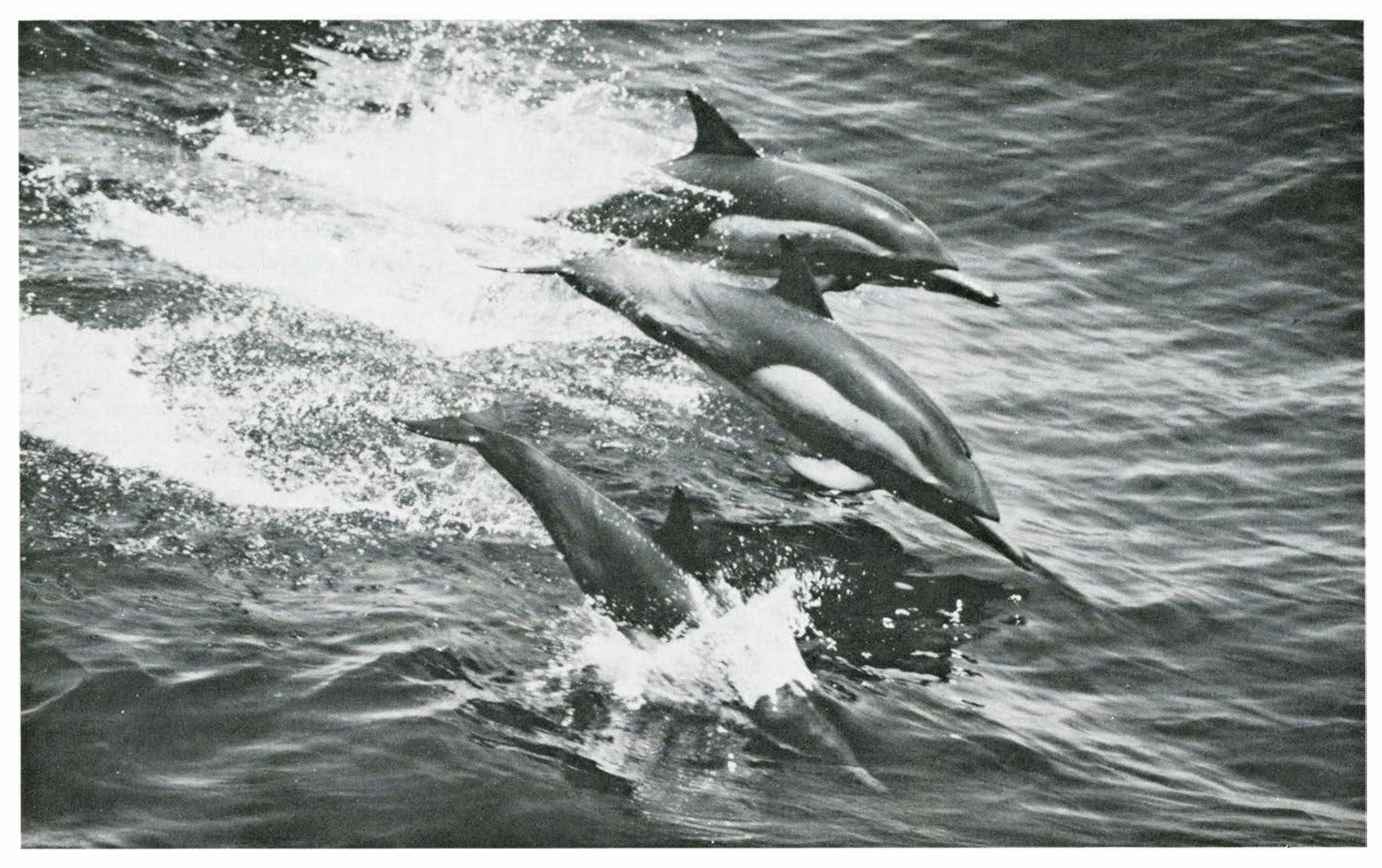
En general, los mamíferos altamente sociales poseen señales sociales complejas y un rico repertorio de comportamiento; de ahí que puedan interactuar con otros miembros de su grupo de maneras muy elaboradas. Entre los ejemplos pueden citarse las señales de agresión que son útiles para establecer y mantener jerarquías de dominancia, las señales de galanteo, los sonidos o movimientos de advertencia cuando se acerca un atacante en potencia y muchas otras señales que contribuyen al funcionamiento del grupo y de cada uno de sus miembros.

El esfuerzo destinado a acumular datos sobre el comportamiento y lo sistemas sociales de los delfines resulta dificultado por el hecho de que la mayor parte de su comunicación se realiza bajo el agua. Es muy difícil acercarse a un grupo de delfines en un bote y permanecer con ellos el tiempo suficiente para empezar a comprender su sistema social. Todo lo que se ve es un grupo de aletas dorsales cuando los animales suben a respirar a la superficie, y después se les pierde de vista cuando se desplazan bajo el agua. Incluso en las raras ocasiones en las que el agua es clara y está en calma, y los delfines pueden observarse algo más que unos pocos minutos, la proximidad del bote puede molestarlos; es, pues, difícil separar lo que es natural en su comportamiento de lo que no lo es y constituye simplemente una reacción frente a la presencia del bote.

Debido a estos y a otros problemas, la mayor parte de las primeras observaciones se hicieron en delfines en cautividad. Aunque la cautividad debe ser una situación anormal para animales que están acostumbrados a una vida con pocas fronteras físicas, los delfines cautivos proporcionaron datos útiles simplemente porque los observadores podían estudiarlos durante períodos prolongados de tiempo. El trabajo mejor conocido de este tipo lo llevó a cabo Margaret C. Tavolga, de Marineland, Florida. Esta investigadora observó un grupo de 12 delfines mulares (*Tursiops truncatus*) en una gran piscina, durante un período total de unos cinco años.

Tavolga encontró que el grupo tenía una jerarquía de dominancia definida. El único macho adulto, que era el más corpulento del grupo, se mostraba más agresivo y menos temeroso que cualquiera de las hembras, de los machos subadultos o de los jóvenes delfines. En general, los animales mayores eran dominantes sobre los de menor tamaño.

Gregory Bateson, del Oceanic Institute de Hawai, obtuvo datos similares. Este autor encontró una jerarquía de do-



TRES FASES DE UN SALTO realizadas (de un modo espontáneo) por tres delfines pertenecientes a la especie *Delphinus delphis* (delfín común) en el mar de Cortés (golfo de California), frente a la costa de México. El repertorio de saltos, giros y volteretas que ejecutan los delfines es variadísimo.

MANADA DE DELFINES de la especie *Delphinus delphis*, el delfín común, fotografiada mientras éstos saltan en el Atlántico. Los delfines son mamíferos sociales que a veces se congregan en manadas bastante grandes, como ocurre aquí, pero que con más frecuencia se encuentran en subgrupos de unos 20 individuos. Son animales de respiración aérea, de modo que sus saltos les permiten en parte respirar. Saltan, asimismo, durante el juego y la caza.

minancia en un grupo compuesto por dos delfines moteados (*Stenella attenuata*) y cinco delfines volteadores (*S. longirostris*). El macho de mayor tamaño amenazaba a los demás delfines (arremetiendo contra ellos o mostrando sus dientes), pero nunca era amenazado. El delfín inmediatamente por debajo en la jerarquía, asimismo un macho, amenazaba a los animales inferiores en jerarquía, y así sucesivamente. Los hallazgos de Bateson demostraron, como también lo han hecho los de otros investigadores, que la jerarquía no es tan estricta como en otros mamíferos. Por ejemplo, el macho al que le correspondía el nivel más bajo podía, no obstante, aparearse con una hembra sin que fuera desafiado por el macho mayor.

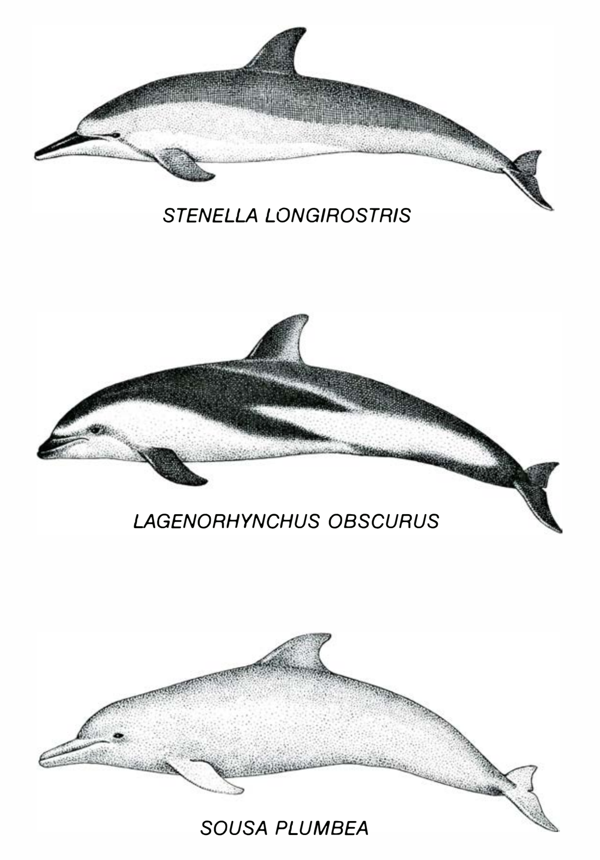
Si la jerarquía de dominancia no se halla necesariamente relacionada con el acceso de los machos a las hembras (como ocurre, por ejemplo, en el harén de los elefantes marinos), cabe preguntarse cuál es su función en las poblaciones de delfines salvajes. Kenneth S. Norris y Thomas P. Dohl, de la Universidad de California en Santa Cruz, han argumentado que la función puede ser organizar a los miembros del grupo para enfrentarse a distintas situaciones. Por ejemplo, las amenazas y las persecuciones por parte de los delfines mayores pueden hacer que las hembras más pequeñas y los animales jóvenes se congreguen en el centro del grupo, donde se hallarán mejor protegidos de depredadores potenciales como tiburones y orcas. Norris cree haber visto esta estructuración en los delfines volteadores y en los delfines moteados.

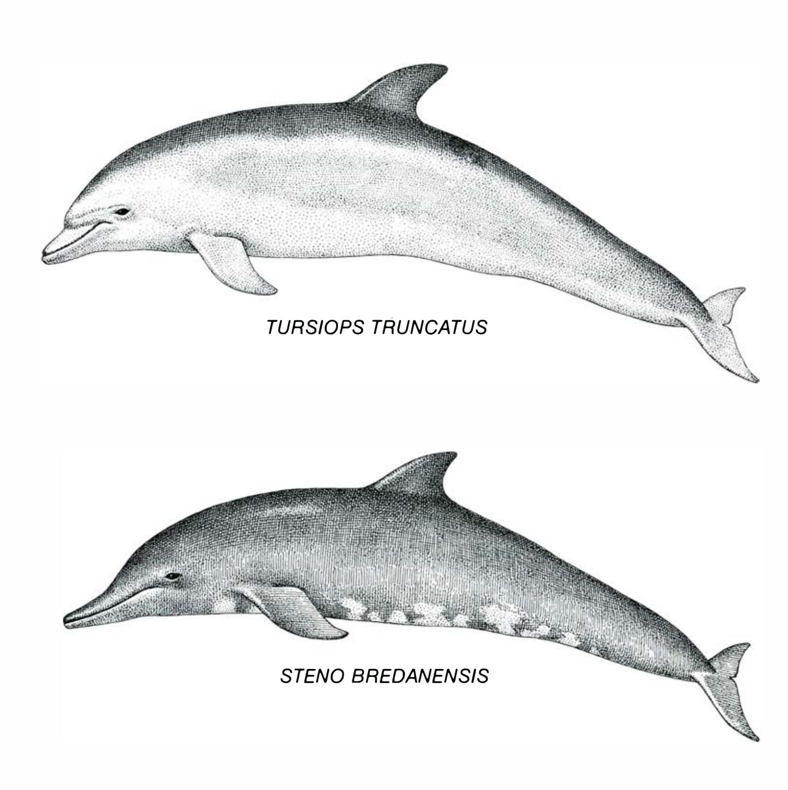
Para que aparezca y se desarrolle un sistema de este tipo es conveniente, y quizá necesario, que los animales que lo poseen se hallen genéticamente emparentados. W. D. Hamilton, del Imperial College of Science and Technology, y otros biólogos han argumentado que los animales estrechamente emparentados tienden a protegerse más entre sí que a proteger a parientes más lejanos, pues los parientes próximos comparten más genes. Si una madre salva a su hijo conduciéndolo al interior del grupo al tiempo que ella misma se pone, al menos parcialmente, en peligro, su aparente altruismo será adaptativo para el grupo porque una proporción significativa de sus genes se preservará.

Si los delfines reúnen de esta manera a los miembros del grupo, es probable que al menos algunos de los animales de un sector del grupo estén emparentados. Asimismo, se ha visto a delfines ayudando a un miembro herido del grupo a alcanzar la superficie para respirar, y protegiendo a un miembro del grupo de los depredadores y de otros peligros. Estas pautas de comportamiento se han citado a menudo como pruebas de altruismo, similar al del hombre, y de gran inteligencia. Parece más probable que representen un subproducto de una tendencia generalizada a ayudar a los individuos emparentados.

Por desgracia, no se conocen los grados de consanguinidad en una población de delfines salvajes. Sólo en fecha reciente se ha conseguido tener alguna idea acerca de la estructura social de los delfines, y este conocimiento está complicado por la existencia de muchas especies distintas y de muchas más poblaciones separadas dentro de cada especie. A partir de los mamíferos terrestres se ha aprendido que el sistema social representa en parte una adaptación al hábitat de la población. Por ejemplo, los lobos que se alimentan sobre todo de ciervos tienden a desplazarse en pequeñas manadas, mientras que los lobos que cazan renos se encuentran en sociedades mayores y mucho más organizadas. Esta diferencia está aparentemente relacionada con la necesidad de un esfuerzo coordinado por parte de varios animales para abatir a la presa mayor.

Recientemente se han llevado a cabo

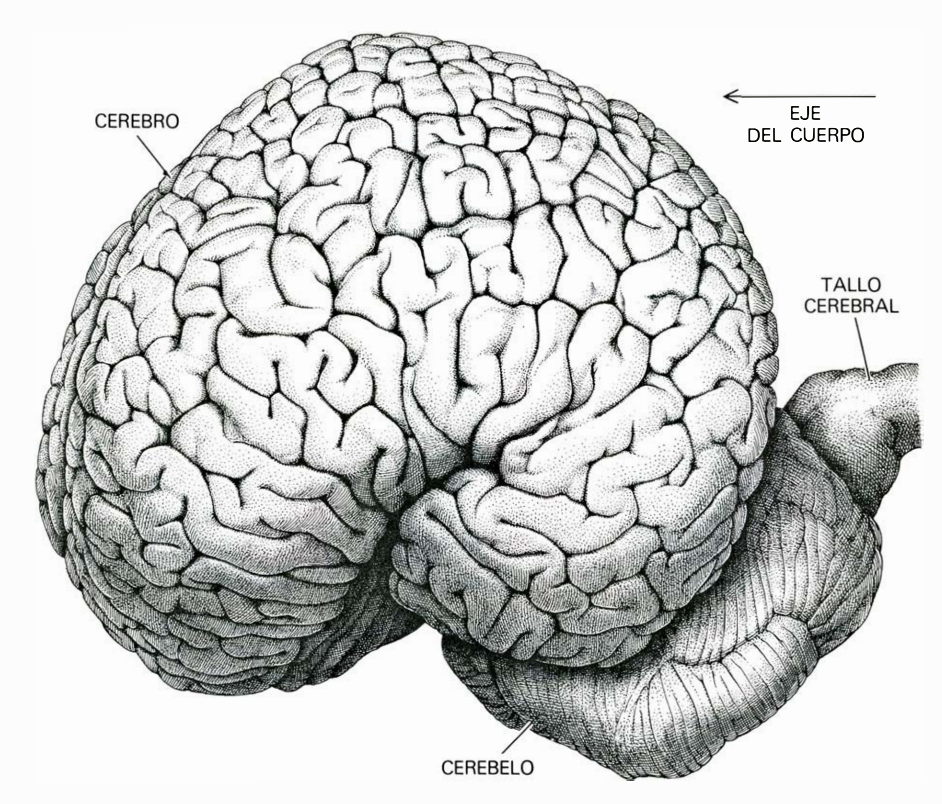




CINCO ESPECIES DE DELFINES ilustradas aquí para indicar sus diferencias en la forma y en la coloración. Se trata del delfín mular (*Tursiops truncatus*), el delfín volteador (*Stenella longirostris*), el delfín obscuro del Atlántico sur (*Lagenorhynchus obscurus*), el delfín jorobado del Indopacífico (*Sousa plumbea*) y el delfín bastardo o de dientes estriados (*Steno bredanensis*). Todos están dibujados a la misma escala. Se conocen alrededor de otras 25 especies de delfines. Las marsopas, bastante similares, tienen un hocico menos puntiagudo y por lo general son más cortas y gruesas.

algunos estudios excelentes que permiten establecer comparaciones entre especies y poblaciones de delfines. Norris y Dohl estudiaron los delfines volteadores hawaianos desde acantilados costeros y bajo el agua. Encontraron que estos mamíferos altamente sociales se desplazaban en bancos cuya media era de 25 miembros. La estructura del banco variaba durante el día de una manera predecible. Por la mañana, los delfines se desplazaban lentamente y en grupos compactos, con los individuos casi tocándose. Parecían hallarse descansando. Andando el día, se volvían cada vez más activos, nadando más deprisa, y algunos de los individuos saltaban fuera del agua al realizar sus volteretas, giros y demás exhibiciones que han hecho famosos a los delfines de los oceanarios.

Por entonces, los bancos se habían hecho más dispersos, y los animales solían hallarse a una distancia de 20 metros unos de otros. Además, los grupos tendían a unirse, de manera que 50 o más miembros podían formar parte del banco ampliado, con todos los animales desplazándose en la misma dirección. Al caer la noche, el banco se desplazaba a varios kilómetros de la costa, entraba en aguas más profundas y los animales empezaban a efectuar inmersiones profundas para alimentarse de peces que capturaban a varios cientos de metros bajo la superficie del agua. El conjunto estaba compuesto por varios grupos muy espaciados situados en el interior de una zona de varios kilómetros de diámetro. Como sea que Norris y Dohl pudieron reconocer a varios individuos, encontraron que un determinado grupo pequeño de animales tendía a mantenerse unido, pero que a menudo los animales cambiaban en bloque de un banco a otro.



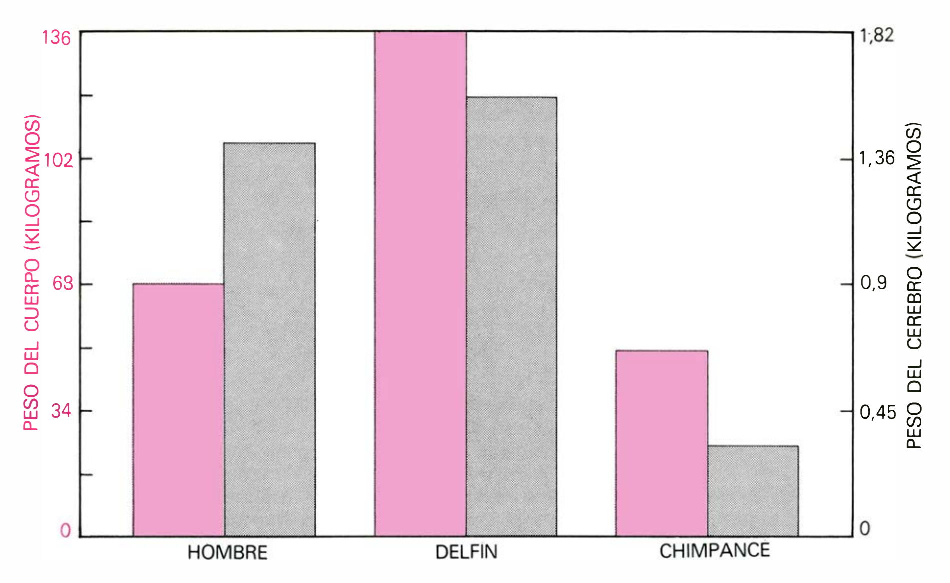
CEREBRO DE UN DELFIN, en este caso de la especie *Tursiops truncatus*, el delfín mular (*véase figura anterior*). El cerebro de un delfín pesa alrededor de 1,6 kilogramos; un cerebro humano típico pesa alrededor de 1,35 kilogramos, pero es mayor que el del delfín en proporción al peso del cuerpo.

Norris y Dohl sugirieron que los delfines forman grupos compactos cuando descansan, de manera que pueden emplear las capacidades sensoriales combinadas de todos los individuos del banco para examinar el ambiente y detectar un peligro potencial. Es bien conocido que los delfines pueden explorar las aguas mediante ecolocación a distancias mucho más elevadas de las que permite únicamente la visión. (En la ecolocación un delfín emite sonidos de alta frecuencia en pulsos cortos, de modo parecido a como lo hacen los murciélagos. Los sonidos rebotan en los objetos, y los ecos informan al animal sobre la distancia, el tamaño, la forma e incluso la textura del objeto.) La hipótesis de Norris y Dohl acerca de las capacidades sensoriales combinadas, en particular durante el descanso, se basa en parte en la capacidad de los delfines de obtener información mediante la ecolocación. Es presumible que cada uno de los delfines de un banco estrechamente organizado pueda oír los sonidos de ecolocación que emiten los demás miembros del grupo. Por tanto, y aunque un determinado individuo puede no emitir muchos sonidos, todos los individuos reciben de manera rápida y eficiente mucha información sobre el ambiente. También es probable que un grupo que se encuentra descansando, nade cerca de la costa con el fin de alcanzar aguas someras que no son frecuentadas por los grandes tiburones de aguas más profundas.

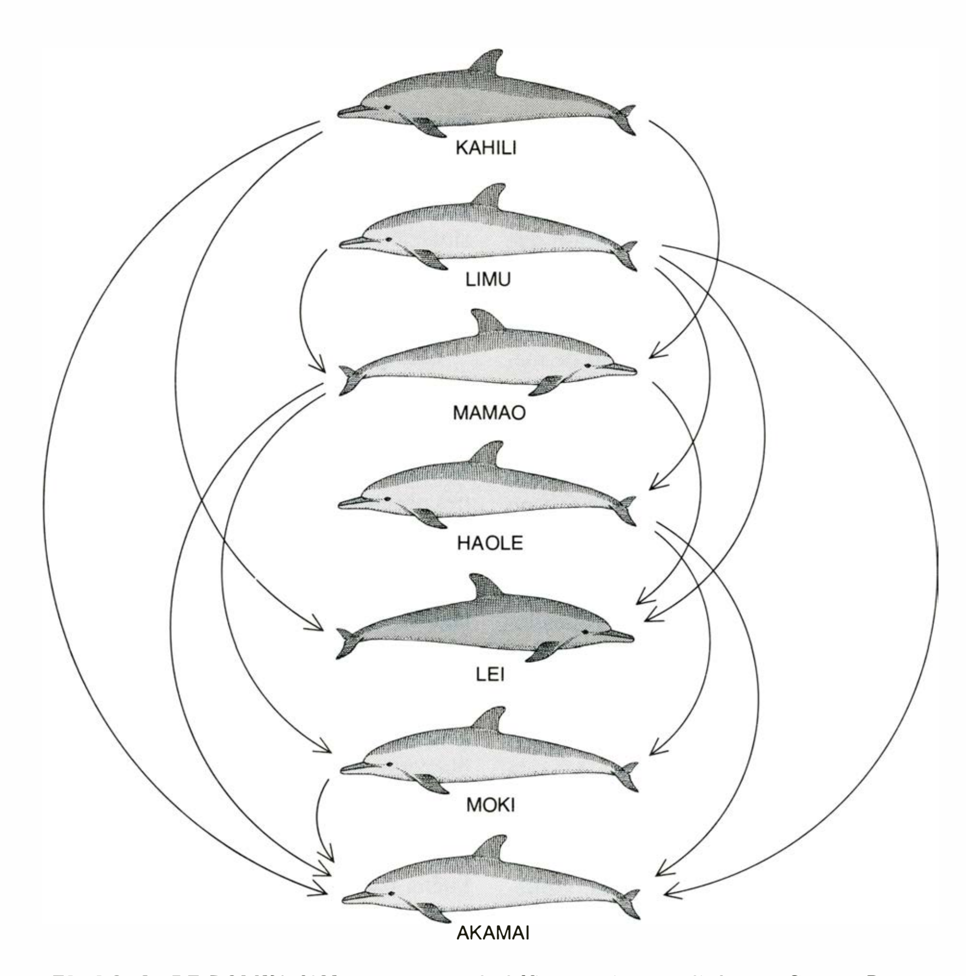
Durante los períodos de alerta, el distanciamiento entre los delfines volteadores aumenta y los animales saltan mucho. Esta actividad puede, en parte, ser un juego, como ha sugerido mucha gente, pero puede representar también una forma de comunicación. A cada salto suele seguir un chapoteo o un golpe fuerte cuando el animal penetra en el agua. Estos sonidos recorren bajo el agua distancias bastante grandes, y pueden informar a otros de la presencia del saltador. De hecho, existen algunas especies de delfines en las que hay grupos que a veces convergen en un cardumen activo y saltarín desde varios kilómetros de distancia. La alimentación durante la noche en aguas profundas está acompañada asimismo por saltos y fuertes chapoteos. En este momento, los miembros de un grupo se hallan muy separados, y en este caso también se supone que los saltos pueden servir para comunicar la situación y quizá también informar, verbigracia, del número de otros delfines que se hallan cerca y de lo que están haciendo.

Graham Saayman y C. K. Tayler, del Port Elizabeth Museum, de Sudáfrica, estudiaron delfines mulares del océano Indico (*Tursiops aduncus*), y mi mujer Melany Würsig y yo estudiamos delfines obscuros (*Lagenorhynchus obscurus*) del Atlántico sur. Ambas especies tienen hábitats similares al del delfín volteador hawaiano. Los hábitats son costero-pelágicos, lo que significa que con frecuencia las tres poblaciones pueden verse y estudiarse desde la costa, pero indica también que se desplazan lejos de ella, generalmente para comer. Las tres poblaciones exhibieron pautas de comportamiento y de movimiento similares. Parece razonable, por ello, decir que el hábitat de estos mamíferos marinos es en buena parte responsable de su modo de vida.

Esta afirmación puede examinarse estudiando delfines en un ambiente distinto. Pueden encontrarse tres de estos ambientes: oceánico profundo, costero y dulceacuícola. Las poblaciones mejor examinadas han sido las costeras. Susan H. Shane, de la Texas A & M Univer-



PESO DEL CEREBRO Y PESO DEL CUERPO comparados para el hombre, el delfín y el chimpancé. El cerebro del hombre pesa alrededor del 2 por ciento del peso del cuerpo, el del chimpancé menos del 1 por ciento y el del delfín algo más del 1 por ciento. Aunque con frecuencia puede obtenerse una idea aproximada sobre la inteligencia comparando la relación peso del cerebro/peso del cuerpo, deben tomarse en consideración otros factores, como la longitud del cuerpo, el modo como se utilizan las extremidades y la complejidad del cerebro. No es posible afirmar, sobre la base de una simple comparación entre el peso del cerebro y el del cuerpo, que un tipo de organismo es más inteligente que otro.



JERARQUIA DE DOMINANCIA en un grupo de delfines cautivos estudiados por Gregory Bateson en el Oceanic Institute de Hawai. El grupo estaba formado por dos delfines moteados (*gris obscuro*) y cinco delfines volteadores (*gris claro*). En este dibujo los machos miran hacia la izquierda y las hembras hacia la derecha. Las flechas indican amenazas (la amenaza supone mostrar los dientes o arremeter) que un animal hace a otro. El macho mayor, Kahili, y el macho Limu nunca eran amenazados. Aunque Akamai, el delfín macho más pequeño situado en el peldaño más bajo de la jerarquía, no amenazó nunca a ningún otro miembro del grupo, sí copuló con la hembra Lei por lo menos en una ocasión.

sity, observó delfines mulares atlánticos en aguas de Texas; A. Blair Irvine, Randall S. Wells y Michael Scott, de la Universidad de Florida, los observaron frente a las costas de Florida; Melany Würsig y yo los observamos en aguas argentinas, y Saayman y Tayler observaron delfines jorobados indopacíficos (del género *Sousa*) en las costas sudafricanas. También aquí se evidenciaban grandes semejanzas. Pueden resumirse mediante una descripción de nuestro estudio de los delfines mulares en aguas de Argentina.

El delfín mular es una especie costera en muchas partes de su área de distribución mundial, de ahí que se le pueda observar fácilmente desde la costa. Durante un período de 21 meses, mi mujer y yo observamos un banco de delfines mulares que pasaba cerca de la costa (siempre a menos de un kilómetro), en aguas de menos de 40 metros de profundidad. Los estudiamos mediante observaciones desde los acantilados costeros y desde un pequeño bote neumático, registrando sus sonidos bajo el agua, fotografiando sus aletas dorsales con el fin de poder reconocer a los distintos individuos y siguiendo sus movimientos con un teodolito de tránsito situado en la costa. (Mediante un teodolito de tránsito de los utilizados por los topógrafos se puede determinar de manera precisa la situación, el movimiento y la velocidad de animales salvajes. El método, desarrollado por Roger Payne, de la New York Zoological Society, se ha demostrado útil en el estudio de los movimientos costeros de los mamíferos marinos.)

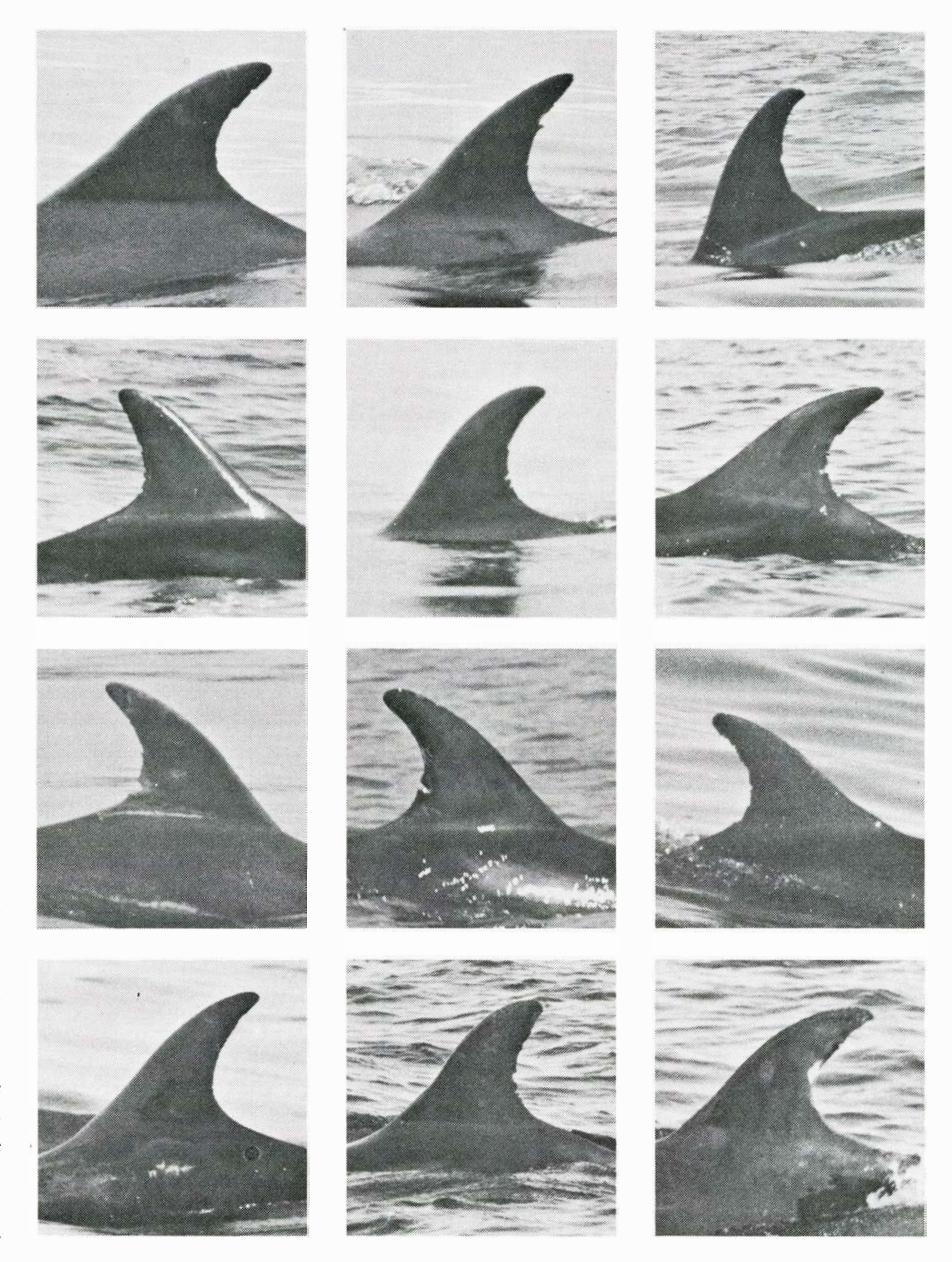
El cardumen variaba en número de ocho a 22 delfines; la media era de 15. Algunos individuos se mantenían continuamente juntos, mientras que otros faltaban en ocasiones. De modo que la población estaba constituida por subunidades de pertenencia variable.

Andando el tiempo, las observaciones iban a corresponder todas a animales que habíamos identificado con anterioridad. Llegamos a la conclusión de que conocíamos a todos, o a casi todos, los aproximadamente 60 delfines que en un momento u otro formaban parte del grupo. Aunque todos los animales del grupo mayor interactuaban en algún grado, nunca vimos en un momento dado a todo el grupo completo. Por tanto, es probable que otros subgrupos, compuestos asimismo por individuos que nos eran conocidos, se desplazaran cerca de otras costas. Al parecer cubren grandes distancias; una vez avistamos un subgrupo de animales conocidos a

más de 300 kilómetros al sur de nuestra base. Es probable que el tamaño de un subgrupo represente un equilibrio óptimo entre el número de animales que se necesitan para la defensa contra los depredadores, el número adecuado para conseguir una alimentación eficiente y el número preciso para la interacción social, la reproducción y la supervivencia de los jóvenes.

En otoño, invierno y primavera los delfines mulares parecían alimentarse de cardúmenes de anchoas hacia mediodía, en aguas de 15 a 35 metros de profundidad. En estas ocasiones avanzaban en un banco disperso, hallándose cada animal a una distancia de unos 25 metros del más cercano. Después de mantener esta formación durante varios minutos, empezaban a bucear y a arremolinarse en una zona. Comparando este comportamiento con el de los delfines obscuros, que en esta región se alimentan casi exclusivamente de anchoas, dedujimos que los delfines mulares se dedicaban a conducir a la superficie del océano a los cardúmenes de anchoas, lugar en el que se las comían. La formación dispersa antes de comer se hace probablemente con el fin de examinar acústicamente una zona tan grande como sea posible en busca de las anchoas.

Esta cooperación entre los miembros de un banco es muy distinta de lo que los delfines hacen en verano. Durante esta época del año no hay anchoas en las aguas costeras de Argentina, de modo que los delfines se alimentan sobre todo de grandes peces solitarios que viven entre las rocas cercanas a la costa. En esta época los delfines se desplazan en aguas que tienen de dos a seis metros de profundidad, y se dispersan en una línea más larga que ancha, hallándose cada animal esencialmente a la misma profundidad y tan cerca como le es posible de la costa. En esta formación, cada uno de los delfines se abre camino entre las rocas y husmea en las grietas mientras busca a sus presas.

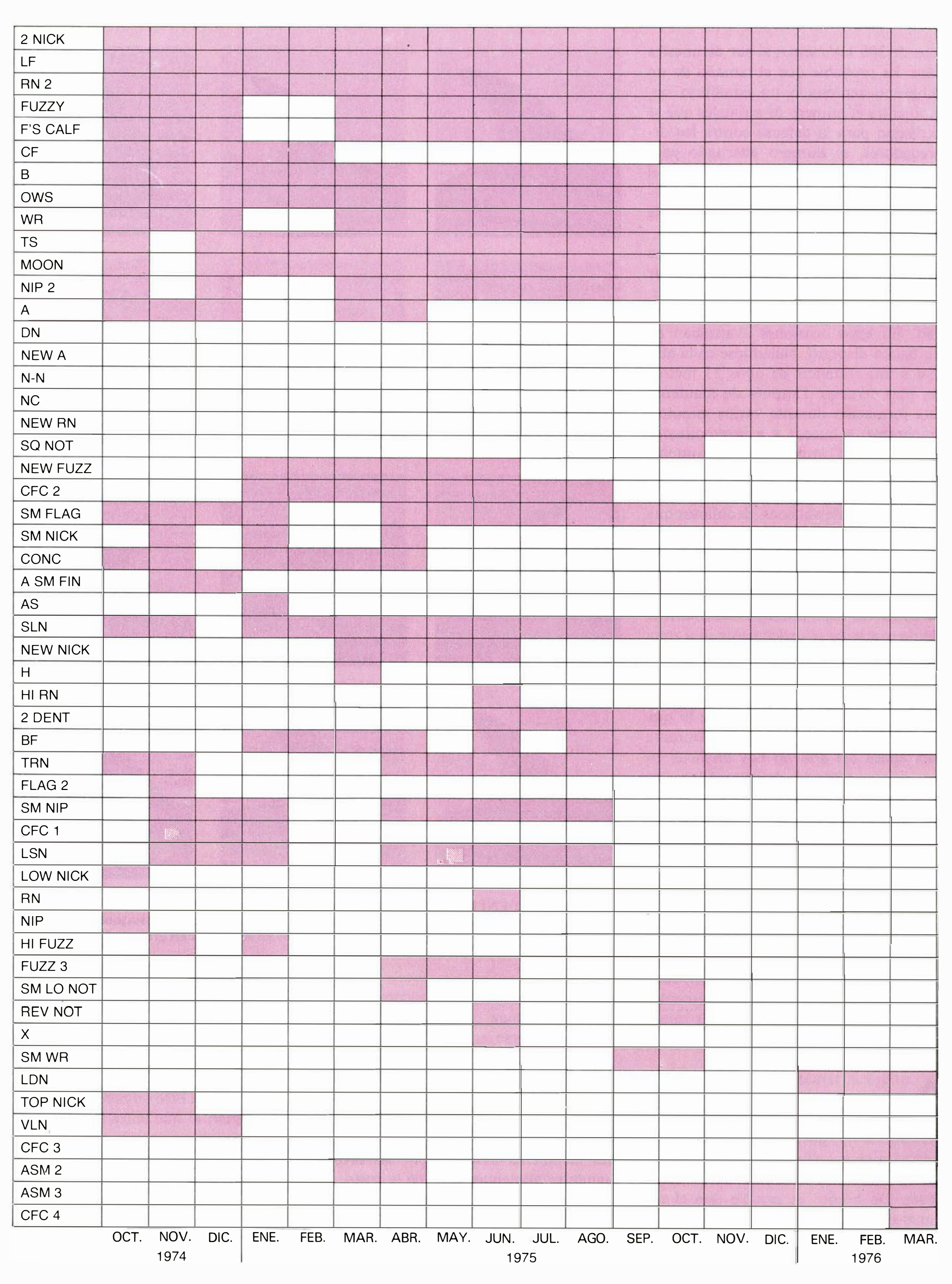


SISTEMA DE IDENTIFICACION de los individuos de un grupo de delfines, basado en el modelo de muescas y cicatrices del borde posterior de la aleta dorsal. Estas fotografías corresponden a una muestra de 12 delfines mulares de un grupo de unos 50. Todos los individuos del grupo podían identificarse.

Aunque el tamaño medio de los subgrupos era de 15 delfines, la media era menor (14) durante el verano que en invierno (20). Puede aducirse que se necesitan menos delfines para una alimentación individual adecuada cerca de la costa. De hecho, es posible que el recurso (grandes peces que habitan en las anfractuosidades de las rocas) se halle limitado, y que un grupo de unos 14 delfines en una zona y en un momento determinados represente el límite superior de la capacidad portadora de aquella zona. El límite es superior cuando los delfines comen de manera cooperativa peces que forman cardúmenes.

Deben tenerse en cuenta, asimismo, otras razones que expliquen la fluctuación estacional en el tamaño de los subgrupos de delfines. Por ejemplo, puede argumentarse que el tamaño de los subgrupos aumenta porque los delfines son más fácilmente atacados por las orcas cuando se alimentan lejos de la costa, y que con más animales por subgrupo pueden protegerse mejor. Quizá las fluctuaciones reflejan máximos estacionales en los apareamientos y en los partos. En base a la información de que disponemos en la actualidad, sin embargo, el alimento parece ser un determinante importante del tamaño (y presumiblemente de la composición) de los subgrupos de delfines mulares.

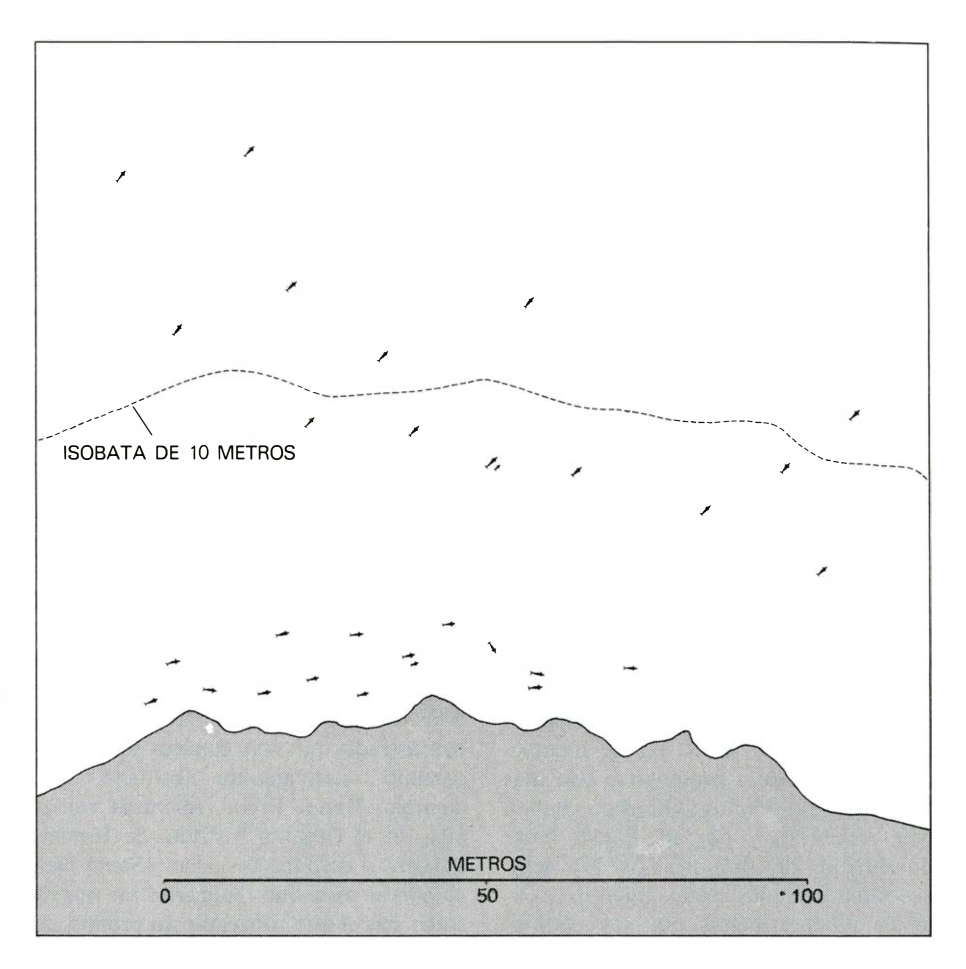
¿Qué puede decirse acerca de la estructura interna de los subgrupos? El hecho de que cada delfín se desplaza de un subgrupo a otro sugiere que los animales tienen lo que entre los mastozoólogos se conoce como una sociedad abierta. Entre los mamíferos terrestres, el chimpancé africano (*Chimpansee troglodytes*) presenta un sistema bastante parecido. Algunos ungulados africanos poseen asimismo un cierto grado de apertura, y los individuos se desplazan frecuentemente de un subgrupo a otro en el interior de una manada definida de manera más rígida. En los chimpancés las variaciones parecen ser una res-



LA COHESION DEL GRUPO en los delfines se indica en este diagrama de la presencia y ausencia de 53 delfines mulares conocidos en aguas de la costa argentina durante un período de 18 meses. Las denominaciones de la izquierda son nombres abreviados que los observadores dieron a los delfines. Una barra coloreada junto a una denominación indica que el individuo fue visto cerca de la costa por lo menos una vez durante el mes correspondiente; un espacio vacío significa que el animal no fue visto durante ese mes. Un mismo delfín pudo haber sido visto más de una vez en aquel mes.

puesta a la disponibilidad de alimento; los animales buscan alimento en pequeñas unidades, pero se congregan en unidades mayores cuando ya han encontrado la comida. Melany Würsig y yo hemos descubierto una situación similar en los delfines obscuros, pero todavía no está claro hasta qué punto una distribución irregular del alimento puede gobernar las variaciones de tamaño de los bancos de delfines mulares.

A pesar de la tendencia a presentar estas sociedades abiertas, casi siempre se observaban grupos de cinco o seis delfines desplazándose juntos. Uno de estos grupos contenía a un animal muy grande y a un segundo adulto que se desplazaron con una cría durante todo el período de estudio de 21 meses. Por ello, creo que el segundo adulto era una hembra. El delfín grande era el único animal que batía regularmente su cola contra la superficie del agua cuando se acercaba un bote. Los amaestradores experimentados de delfines dicen que el golpear con la cola puede significar cólera, agresión o advertencia. Sería tentador decir que el delfín grande era en cierto sentido el líder del grupo, pero no tengo información suficiente acerca del sistema social y del comportamiento de los delfines para verificar esta hipótesis.



FORMACIONES DE GRUPOS de delfines mulares frente a las costas de Argentina. Variaban según los animales se hallaran en aguas someras cerca de la costa o en aguas profundas más alejadas de ella. En aguas someras, cazaban individualmente peces que habitaban entre las rocas, mientras que en aguas profundas actuaban en grupo para descubrir cardúmenes de anchoas y conducirlos a la superficie.

Otros dos grupos de cinco y seis delfines que siempre encontramos juntos estaban formados por adultos de aproximadamente el mismo tamaño. Es posible (también aquí estoy haciendo suposiciones) que los grupos estuvieran formados por miembros no reproductores de la población, de manera parecida a como los elefantes solteros viajan juntos en manadas.

Los descubrimientos más detallados sobre la composición por edad y sexo de los bancos de delfines proceden de los estudios de Irvine, Wells y Scott sobre los delfines mulares de Florida. Capturaron 47 ejemplares y les colocaron etiquetas y otras marcas de identificación con el fin de poderlos reconocer más tarde en su hábitat natural. Mientras marcaban a los animales pudieron determinar el sexo y el tamaño de los individuos.

Después de soltarlos, Irvine, Wells y Scott encontraron que el área vital de la manada residente cubría unos 85 kilómetros cuadrados. Con frecuencia, las hembras y las crías se desplazaban en grupos que incluían a unos pocos machos adultos o a ninguno. Estos machos tendían a asociarse más frecuentemente con hembras sin crías que con madres y jóvenes, y raramente se asociaban con machos subadultos. Los machos subadultos se encontraban en ocasiones en grupos de solteros alejados de los demás delfines. Algunas hembras fueron observadas en compañía de sus crías durante un período de hasta 15 meses. De ahí se deduce que existe un fuerte lazo social entre la madre y la cría, que continúa probablemente mucho después del destete. No se ha observado una asociación a largo plazo de este tipo entre un macho y una cría.

Debe destacarse que las relaciones sociales de los delfines no son estrictas ni inmutables; por el contrario, son muy variables. No obstante, existen unas cuantas características principales. Los lazos entre madre y cría son fuertes; los lazos entre macho y hembra y entre macho y cría parecen serlo menos. Esta comparación sugiere que el apareamiento es relativamente promiscuo. Los machos subadultos pueden verse excluidos de la rutina social normal, pero las hembras subadultas no lo están, lo que sugiere que los machos adultos pueden copular más con varias hembras que las hembras con machos distintos. Una relación de este tipo representa un indicio de sistema de apareamiento poligínico, que también es común entre los mamíferos terrestres.

Puede decirse menos acerca del comportamiento de los delfines que de sus relaciones sociales, aunque ha ido acumulándose un volumen considerable de datos de comportamiento a partir de las observaciones de individuos en la naturaleza y en cautividad. Creo que todavía no se sabe lo suficiente para apoyar ninguna afirmación definitiva y amplia. Sin embargo, pueden citarse unos cuantos ejemplos interesantes de comportamiento. Los delfines mulares (y otras especies) parecen realizar galanteos y cópulas a lo largo de todo el año, lo que en la naturaleza se advierte frecuentemente cuando dos animales nadan juntos vientre contra vientre. No obstante, los delfines mulares y algunas otras especies presentan un claro máximo de alumbramientos anual (a veces dos máximos). En los delfines que observamos frente a las costas de Argentina, todas las crías nacieron en verano. Este descubrimiento indica que un cambio fisiológico en el macho o en la hembra hace que la concepción tenga lugar en un período

limitado. Este cambio se ha documentado en forma del aumento estacional del peso de los testículos en los machos de varias especies de delfines.

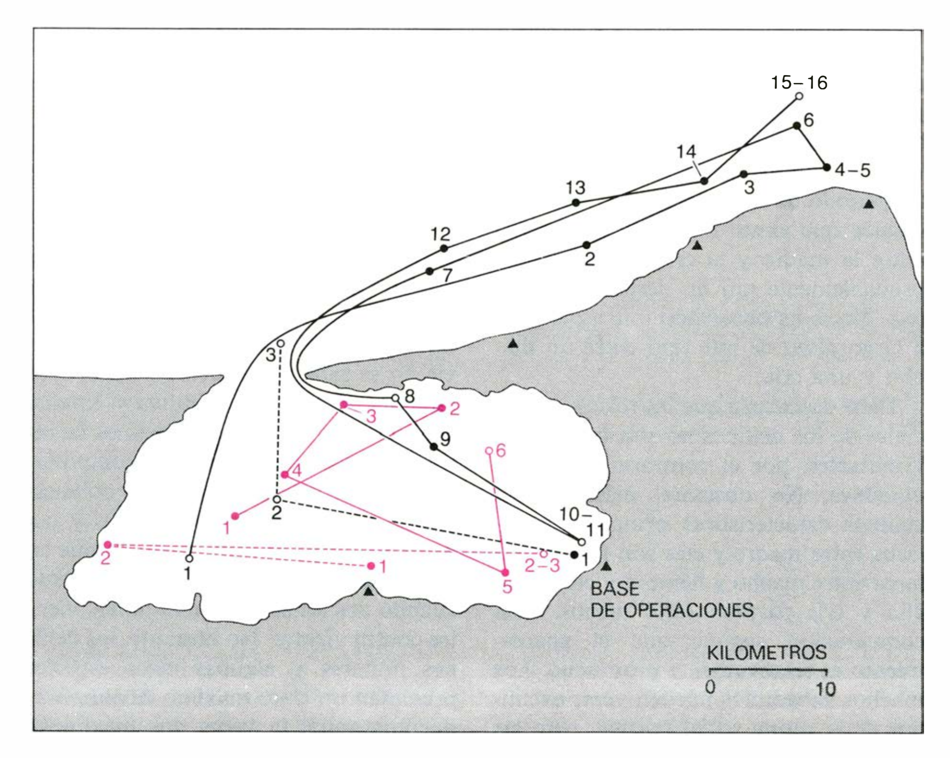
Un apareamiento realizado a lo largo de todo el año implica además que el cortejo puede tener otras connotaciones además de la sexual. Varios investigadores han sugerido que estas interacciones pueden servir asimismo para definir y reforzar las jerarquías y los lazos sociales. El argumento viene apoyado por la frecuente actividad homosexual que se observa por lo menos entre los delfines cautivos. Estudios futuros pueden demostrar que sería más adecuado calificar este comportamiento de "homosocial". En muchos otros grupos de mamíferos se advierte un remanente de señales sexuales que se incorporan a las jerarquías de dominancia.

Un segundo tipo de comportamiento que se observa en casi todas las especies de delfines es el salto. Ya he mencionado que tiende a presentarse con más frecuencia cuando los animales se hallan bien separados y por ello puede tener una función de comunicación. Los delfines mulares de las costas argentinas saltaban mucho menos que los delfines obscuros de la misma región, aunque las dos especies cazan peces de una manera esencialmente similar. Sin embargo, los delfines mulares se desplazaban en un cardumen y por ello no tenían demasiada necesidad de comunicarse con otros miembros del banco mediante saltos. Los delfines obscuros se desplazaban en hasta 30 pequeños bancos en un lugar determinado, y a menudo vimos que los cardúmenes se fusionaban cuando empezaba el salto. Sin duda, los saltos tienen otras funciones, como la de ayudar a conducir o a cazar a las presas, pero la función de comunicación entre los miembros de una especie determinada puede ser importante.

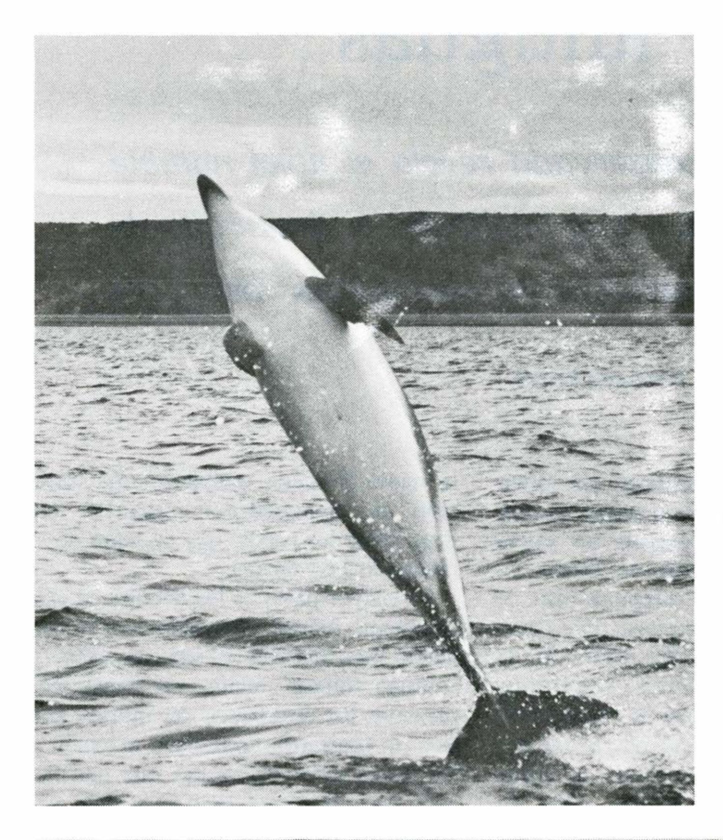
Sobre la base de lo relativamente poco que en la actualidad se sabe acerca de los delfines, ¿qué puede decirse de su inteligencia? Ciertamente, los delfines son diestros en aprender tareas complejas, como lo demuestran con los trucos que realizan en los oceanarios, y recuerdan estas tareas durante años. También se ha demostrado que son capaces de un raciocinio relativamente abstracto. Por ejemplo, Karen Pryor, mientras trabajaba en el Oceanic Institute de Hawai, adiestró a delfines bastardos (*Steno bredanensis*) para que realizaran un nuevo truco con el que conseguir un premio en forma de peces. Después de varios días de adiestramiento, estos delfines realizaban tipos de saltos y de contorsiones completamente distintos, aparentemente al "comprender" que las formas de comportamiento que habían exhibido previamente no iban a ser premiadas. Pero muchos amaestradores han señalado que lo mismo ocurre en los perros y otros mamíferos, y posiblemente quizá también en las palomas, lo que implica que no hay necesidad de postular una superinteligencia para explicar qué hacen los delfines.

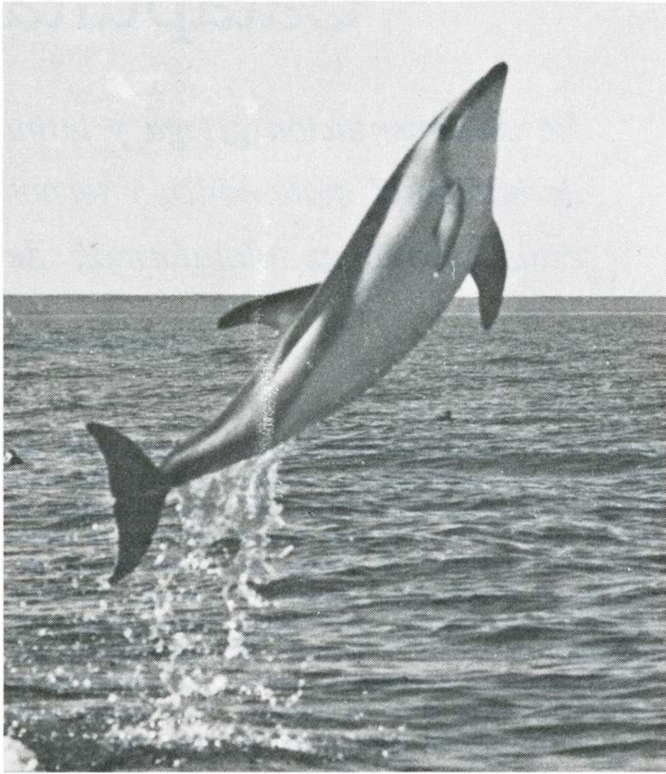
Edward O. Wilson, de la Universidad de Harvard, ha sugerido que el cerebro del delfín puede ser mayor en relación al tamaño y al peso del cuerpo que el cerebro de la mayor parte de los demás mamíferos, debido precisamente a la famosa capacidad de imitación que ha hecho de los delfines los animales favoritos de los adiestradores y del público. La pregunta que se plantea entonces es la siguiente: ¿Por qué razón ha de ser un animal un imitador tan fenomenal? R. J. Andrew, de la Universidad de Sussex, ha hecho notar que el mimetismo vocal puede ser importante en animales que con frecuencia viajan sin poderse ver entre sí. Los miembros de un grupo muy disperso podrían reconocer así a otros miembros debido a una complicada convergencia de señales entre los animales de un grupo o manada. Se ha demostrado que este sistema, en forma de dialectos, funciona en algunos primates y aves. Es plausible que en los delfines, animales que confían mucho más en el sonido, haya evolucionado en mayor grado un sistema semejante.

¿Y qué decir de la imitación del movimiento? Wilson sugiere que los delfines pueden imitar a los miembros del grupo que tienen más éxito a la hora de capturar peces y de evitar ser depredados. Además, resulta ventajoso que los animales que forman parte de sociedades sociales que cooperan para cazar el alimento (lo que se ha demostrado por lo menos en algunas especies de delfines) conozcan bien los movimientos de los otros miembros del grupo, para que los individuos sean capaces de desempeñar varios papeles en el momento de reunir a un cardumen de peces. Wilson señala que la imitación por sí sola es suficiente para explicar el tamaño del cerebro de los delfines, y que las señales sociales de estos cetáceos no son probablemente más refinadas que las de la mayor parte de mamíferos y aves. En mi opinión, no se sabe lo suficiente acerca de las señales sociales de los delfines para que una afirmación de este tipo pueda tener fundamento. Norris cree que los poderes de imitación de los delfines no son necesariamente superiores a los de muchos



A TRAVES DE LA LOCALIZACION POR RADIO de cuatro delfines en aguas de Argentina (en el golfo de San José y en el mar abierto adyacente) se obtuvieron estas pautas de movimiento. Cada delfín está representado por una ruta diferente, y el número junto a cada círculo indica la posición del animal a este número de días después de que se le dotara de un transmisor de radio. Los círculos negros corresponden a posiciones conocidas, obtenidas mediante triangulación desde los puntos costeros indicados por los triángulos o por haberse acercado al delfín en un bote, y los círculos vacíos representan posiciones estimadas. La situación diaria de los animales se ha convenido en que sea la del mediodía.







VARIEDAD DE SALTOS que pueden realizar los delfines; dicho repertorio se advierte a partir de estas fotografías de tres delfines obscuros (*Lagenorhynchus obscurus*) en mar abierto. Se ha observado que, en cautividad, aumentan la variedad en sus saltos cuando se les ofrece premios de comida.

otros mamíferos cuyo cerebro es más pequeño y menos complejo físicamente. En la actualidad parece fútil comparar la inteligencia de los delfines con la de otros mamíferos, simplemente debido a la falta de información apropiada sobre los delfines y a las grandes diferencias entre su ambiente y el de los mamíferos terrestres.

La necesidad de poseer mejor información acerca de los delfines hace que uno piense en mejores métodos de obtenerla. Una posibilidad es intentar habituar a los delfines a los observadores hasta el punto de que aquéllos realicen sus actividades diarias como si los observadores no estuvieran presentes. George Schaller estudió de este modo a los gorilas de montaña, y Jane Goodall abrió de manera parecida una nueva era en la investigación de los chimpancés. Se desplazaban con los animales y se sentaban pacientemente hasta que éstos los aceptaban o simplemente los ignoraban.

¿Cómo puede seguirse, en pleno océano, a un grupo de delfines? Quizá no sea necesario. Jody Solow, de la Universidad de California en Santa Cruz, aprendió recientemente a emitir bajo el agua un sonido que solía atraer hacia ella a los individuos de un grupo de delfines volteadores hawaianos que se hallaba en las cercanías. Su éxito deja abierta la posibilidad de que con el tiempo un investigador pueda reconocer a todos los miembros de un grupo, aprender sus pautas e interacciones sociales y hacerse una idea más cabal de su comportamiento natural.

Catapultas antiguas

La documentación griega y latina conservada revela el nivel notable de habilidad matemática y tecnológica que implicaba el desarrollo de estas primitivas plataformas de lanzamiento de misiles balísticos

Werner Soedel y Vernard Foley

En el año 399 a.C. Dionisio el Antiguo, que gobernaba la colonia griega de Siracusa en Sicilia, preparó su ciudad para una larga guerra con Cartago, poniendo en marcha un programa de investigación y desarrollo. Utilizó técnicas que, hoy, nos resultan familiares, tales como la reunión de grandes equipos de especialistas, la división del trabajo con el fin de distribuir las tareas en unidades manejables y la creación de incentivos psicológicos y financieros. Con ello, Dionisio se orientó claramente hacia la producción de armas nuevas. Resultados del programa fueron los cuatrirremes y, posiblemente, los quinquirremes, naves con cuatro o cinco bancos de remos que tenían, por tanto, más energía potencial detrás de sus espolones que los trirremes de tres bancos. Los ingenieros de Dionisio diseñaron también las prìmeras catapultas.

Estas máquinas primitivas lanzaban probablemente flechas con un arco que no debía ser mucho más poderoso que el que un hombre podía tender. No obstante, al mecanizar el proceso de tensión y liberación de la flecha, los inventores de las catapultas hicieron posible la construcción de arcos mucho más potentes. Parece que estos artilugios se construían con materiales compuestos, con un núcleo de madera coronado por una capa de tensión fabricada con tendones animales por la parte de delante y con otra capa de compresión, hecha de cuerno, por detrás. Con el tiempo, el arco flexible alcanzó el límite de las posibilidades de su diseño y fue sustituido por catapultas basadas en el principio de torsión. El nuevo sistema utilizaba ovillos muy tirantes de fibras elásticas que recibían un refuerzo suplementario, cuando el arma estaba plenamente tensada, mediante un brazo rígido de arco. Podía utilizarse crin de caballo o cabello humano para las cuerdas que formaban estos ovillos pero, para obtener mejores rendimientos, se prefería el uso de tendones animales.

Con el fin de mecanizar los movimientos del arquero, los ingenieros de las catapultas incorporaron un cierto número de características de diseño. La pieza básica de la catapulta era el mango, una barra compuesta que constituía el eje principal del arma. En la parte superior del mismo se encontraba una ranura sobre la que se ensamblaba una segunda barra, la corredera, que podía moverse hacia adelante y hacia atrás. Esta corredera llevaba, sobre su superficie trasera, un mecanismo a base de retén y gatillo, que permitía sujetar y liberar la cuerda del retén. Sobre la corredera, frente al retén, se encontraba una acanaladura en la que se depositaba la flecha y sobre la que se efectuaba el lanzamiento de la misma. En el momento de la acción, la corredera se movía hacia adelante hasta que el retén sujetaba la cuerda del arco. Luego se forzaba la corredera, que llevaba consigo la cuerda, hacia atrás hasta que el arco alcanzaba su máxima tensión. En las versiones primitivas, retenes lineales, situados en la parte posterior del mango, se engarzaban con lengüetas de la corredera para resistir la fuerza del arco. Más tarde se adoptó un retén circular situado en la parte posterior del mango. En las primeras catapultas se forzaba a mano el retroceso de la corredera, pero pronto el tamaño y potencia de las máquinas requirió el uso de una manivela.

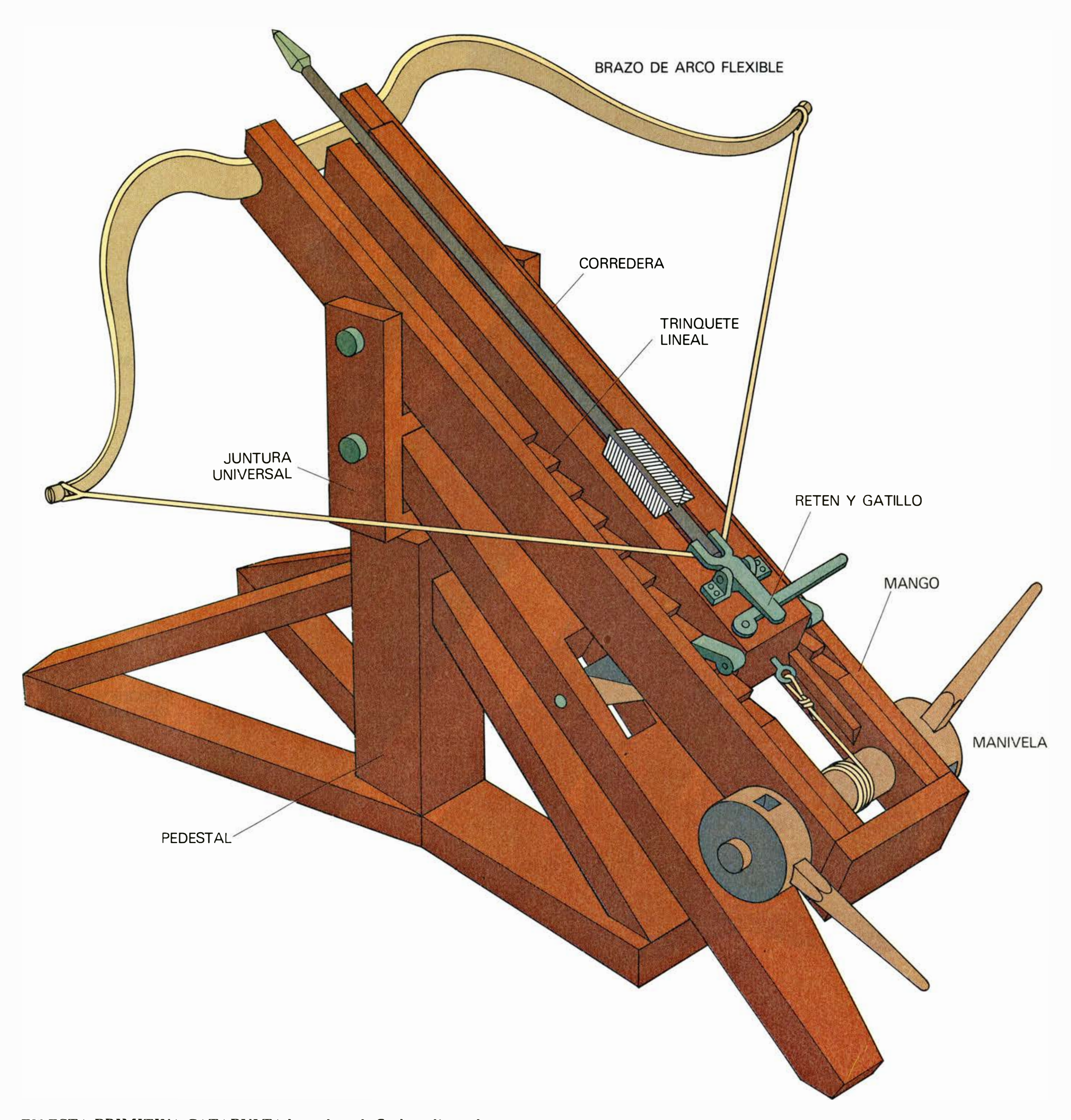
Cuando la catapulta resultó demasido grande para ser disparada sobre el hombro, se la colocó sobre un pedestal. Para que resultara más fácil el apuntar con ella se inventó una juntura especial que conectaba el mango con el pedestal. La solución de este problema anticipó el invento de la juntura universal que se atribuye generalmente a Girolamo Cardano o a Robert Hooke y, por consiguiente, a los siglos XVI o XVII. Del mismo modo las ensambladuras a cola de pato de la corredera aparecen de manera general en la construcción de máquinas modernas y el uso de un retén que substituya a la mano humana estableció una tradición de manipulación mecánica que llega hasta los robots que se están desarrollando hoy en día.

La mayor parte de los detalles de esta tecnología fueron silenciados por la tradición más literaria del mundo griego y romano antiguo. Incluso en etapas más recientes los estudiosos del mundo clásico no prestaron demasiada atención a los textos sobre catapultas que han llegado hasta nosotros. Sólo en el siglo XX, eruditos que asociaban una formación de ingenieros con experiencia militar empezaron a descifrar los antiguos tratados sobre catapultas y entonces se puso de relieve su importancia. El oficial de la artillería alemana Erwin Schramm hizo labor de pionero con estudios que llevaron a la construcción de armas reales, elaboradas de acuerdo con estos textos, poco después de los comienzos de este siglo, y esto fomentó una línea de investigación que ha culminado en los últimos años con los trabajos definitivos del historiador británico E.W. Marsden. Esta literatura ha puesto de manifiesto que los griegos no desdeñaban, tal como se ha supuesto a menudo, la observación ajustada y el experimento exigente. Tal vez Platón despreció el fracaso de los objetos del mundo real que no lograban ajustarse a las dimensiones ideales de la geometría, y quizás Aristóteles basó, en buena parte, su biología en descripciones meramente verbales de las especies, pero en el seno de la comunidad de antiguos ingenieros mecánicos se desarrollaban métodos para controlar la naturaleza que resultaron mucho más importantes cara al futuro.

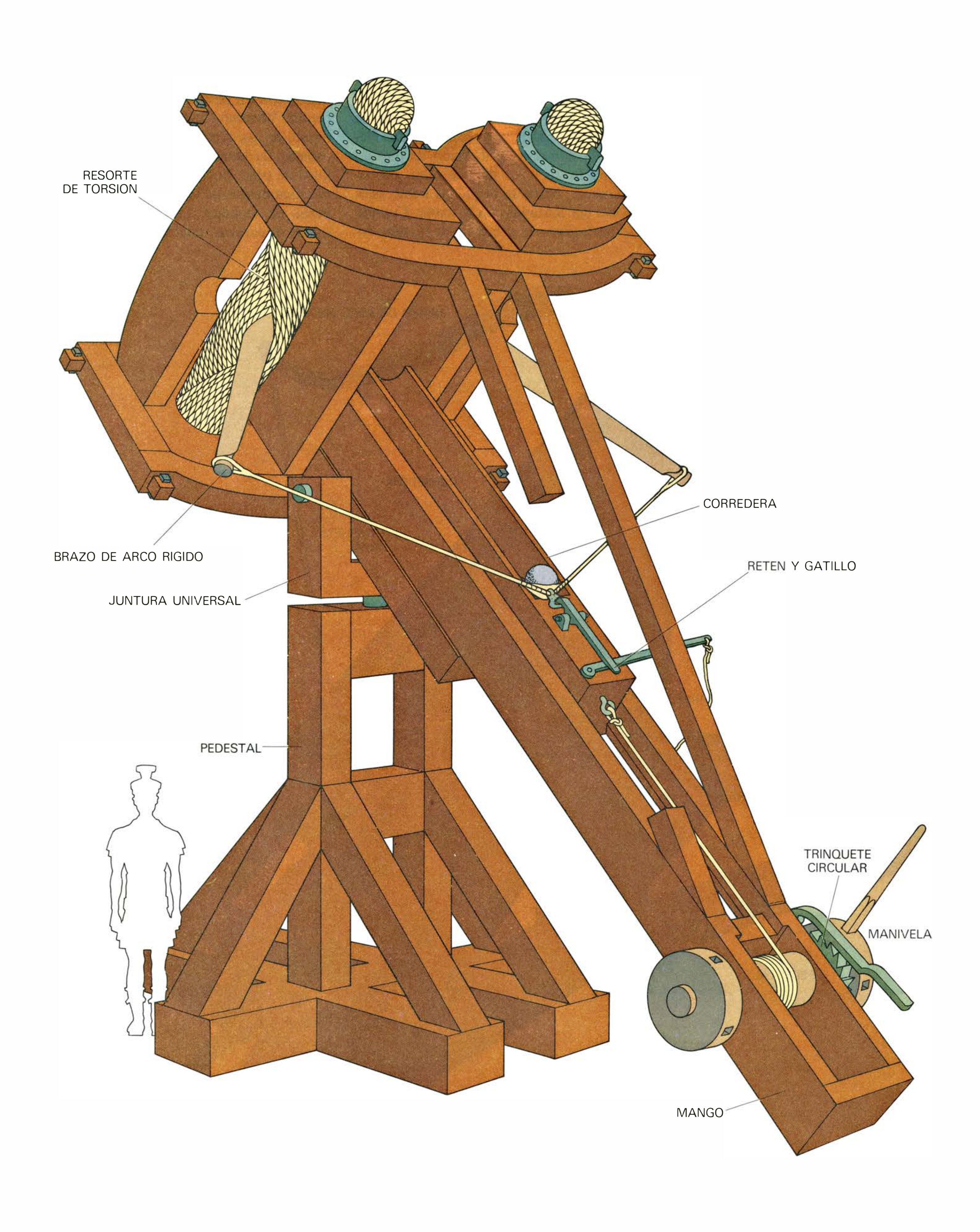
La sustitución del arco flexible por el resorte de torsión supuso un enorme avance para la ingeniería de las catapultas.Este salto hacia adelante se realizó, aproximadamente, medio siglo después de la invención de la catapulta y hay pruebas de que Filipo de Macedonia, padre de Alejandro Magno, fue el gober-

nante que financió esta nueva etapa de investigación y desarrollo. Catapultas muy poderosas aparecieron por primera vez con ocasión de las campañas de Alejandro. Desde luego tenemos motivos para asociar la aparición de grandes imperios con el surgimiento de la catapulta.

La catapulta basada en el arco flexible se limitaba a lanzar flechas o piedras relativamente pequeñas, y las piedras requerían una cuerda de arco provista de una bolsa en su parte central. Las flechas debían tener el tamaño de una jabalina ligera y las piedras eran, probablemente, bastante pequeñas como para poder ser lanzadas con la mano. Cuando se perfeccionó el principio de torsión, resultó perfectamente posible lanzar una piedra con un peso de hasta 78 kilos. De hecho Vitrubio, ingeniero militar romano, da las dimensiones de catapultas capaces de lanzar piedras de hasta 162 kilos de peso, por más que tales máquinas gigantescas tal vez nunca se construyeron en realidad. Las máquinas habituales disparaban piedras comprendidas



EN ESTA PRIMITIVA CATAPULTA lanzadora de flechas, diseñada por técnicos que trabajaron para Dionisio el Antiguo de Siracusa en el siglo IV a.C., el arco flexible se montaba al extremo de una larga armazón de madera que abarcaba una corredera a cola de pato. La corredera móvil, que arrastraba consigo la cuerda del arco mediante un dispositivo a base de retén y gatillo, se sujetaba a la parte posterior del mango, manteniendo el arco en tensión, mediante un trinquete (retén) lineal, después de tensarlo con ayuda de una manivela. La pieza que conecta la catapulta a su pedestal parece haber sido una versión antigua de la juntura universal. El mismo arco constaba, probablemente, de tres materiales distintos encolados juntos: un núcleo de madera, una capa anterior hecha con tendón animal y otra posterior de cuerno. Dado que el tendón es fuerte en tensión y el cuerno en compresión, tales arcos resultarían mucho más potentes que los corrientes, construidos con una sola pieza de madera. La flecha mide unos dos metros.



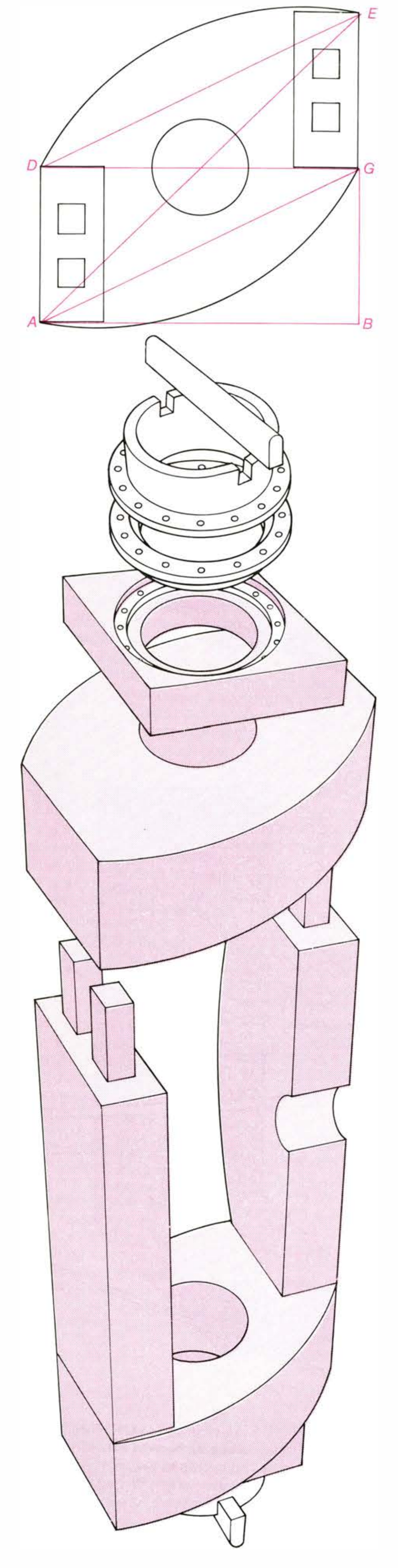
LOS RESORTES DE TORSION permitieron a los antiguos ingenieros de catapultas diseñar armas mucho mayores como esta versión romana de la catapulta lanzapiedras, capaz de disparar un proyectil que pesara un talento, o sea, 26 kilos (un dibujo frontal de un artilugio similar aparece en la portada). Una bolsa, sujeta a la parte central de la cuerda del arco, retiene la piedra y un anillo unido a la parte posterior de la bolsa engarza con el gatillo. Las mordazas situadas en los extremos de los ovillos de torsión podían girarse y luego sujetarse para ajustar la tensión antes de disparar.

entre 13 y 26 kilos. Aparecen, ahora, máquinas que disparan flechas de casi cuatro metros de longitud. Es evidente que el arquero o el hondero no podía competir en alcance ni siquiera con los proyectiles lanzados por catapultas de escala más reducida. El máximo alcance documentado para una catapulta que disparara una flecha de tamaño ordinario, de unos 70 centímetros, es de 640 metros aproximadamente y hay motivos para creer que tal pretensión no es exagerada. El alcance máximo de un arquero era de unos 450 metros.

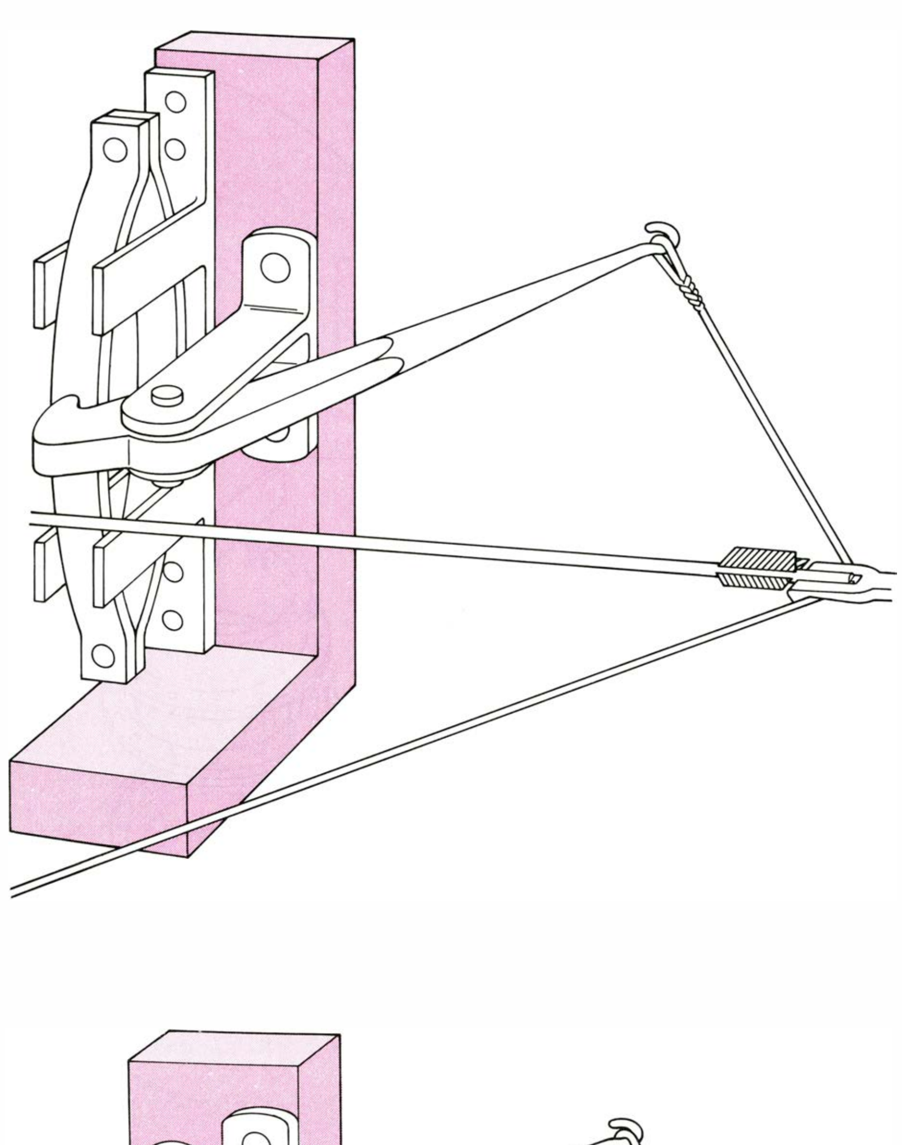
Las catapultas podían lanzar tales proyectiles con una precisión notable. Podía fácilmente concentrarse su fuego sobre un solo blanco mediante disparos sucesivos, con lo que se lograba derribar las almenas situadas sobre las murallas de una ciudad o desgarrar las armaduras protectoras de una torre móvil de asedio. Se podían apuntar las catapultas durante el día, cuando eran observables los impactos de los proyectiles, y aterrorizar más tarde al enemigo disparando a intervalos durante la noche sobre el mismo blanco. En el sitio de Avaricum, en el 52 a. de C., Julio César observó que sus catapultas derribaban con facilidad a los guerreros galos uno tras otro, ya que éstos debían situarse en una posición muy arriesgada que resultaba vital para el avance del ataque romano. Todos estos detalles apuntan a un alto grado de precisión en el tiro. (De hecho cuando una de las catapultas construidas por Schramm fue puesta a prueba, en presencia del Kaiser, partió en dos, de acuerdo con los informes oficiales, una flecha con la flecha siguiente, en línea con la mejor tradición de Robin Hood.) Evidentemente, la mecanización había superado con mucho las posibilidades de los arqueros humanos. Los resultados debieron ser sobrecogedores para los que tenían que enfrentarse con tales máquinas.

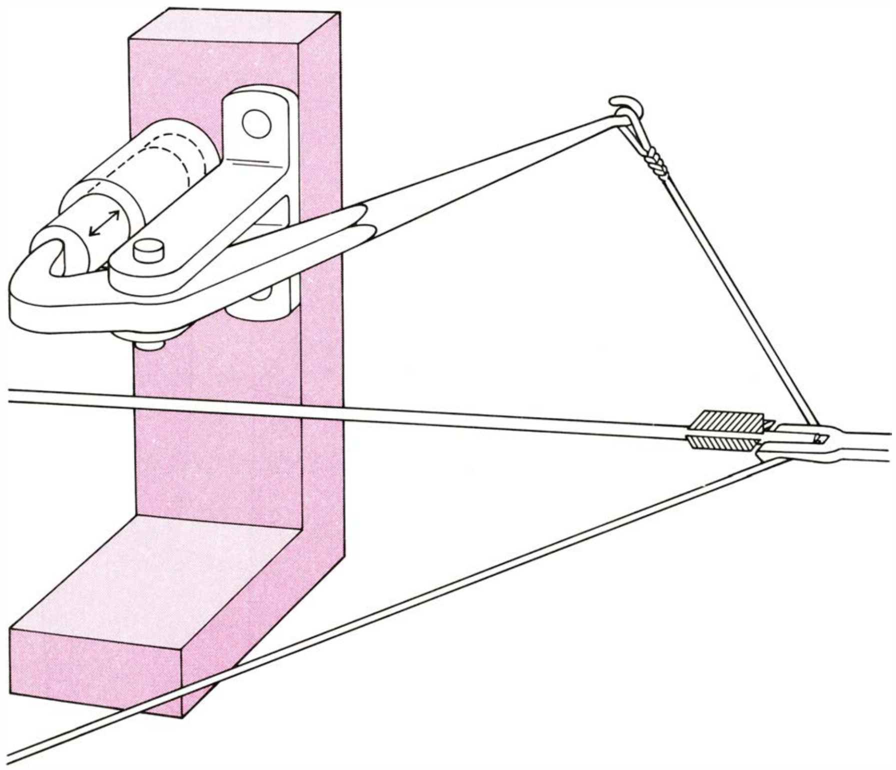
Durante el sitio de Jerusalén en el 63 a. C., Josefo, el comandante de las fuerzas judías que defendían la ciudad, menciona que la cabeza de un amigo que se encontraba a su lado fue arrancada de cuajo por un proyectil lanzado por una catapulta romana. Incluso a distancias de casi 400 metros uno de esos proyectiles podía, aparentemente, atravesar varias filas de soldados antes de detenerse por completo. En el mismo asedio, según Josefo, una mujer embarazada fue muerta por un proyectil y el feto salió lanzado a una distancia de más de treinta metros. Las flechas largas y pesadas resultaban igualmente eficaces: su caída podía clavar, literalmente, en el suelo a las tropas que avanzaban. De hecho, la palabra catapulta deriva de la capacidad de penetración de la máquina. Las tropas griegas de armamento ligero llevaban un escudo llamado *peltē* y el prefijo *kata* denota movimiento descendente. Por tanto, una catapulta es un instrumento capaz de destrozar por completo, en su movimiento descendente, a través de un escudo. En el sitio de Gaza en el 332 a.C., Alejandro fue herido en el cuello por una flecha, lanzada por una catapulta, que había atravesado tanto su escudo como su coraza. Se sabe que las máquinas de Arquímedes inspiraron terror durante el sitio de Siracusa por los romanos entre 213 y 211 a.C. Una típica reacción de la época es la de un general espartano, Arquidamus, quien, al contemplar cómo se disparaba una catapulta, exclamó proféticamente: "¡Oh Hércules, el valor marcial de los hombres ya no sirve para nada!".

En resumen, las catapultas alteraron de manera significativa la dirección de las guerras así como el equilibrio de la política y de la sociedad. En términos



DIMENSIONES DE UN TABLERO que constituía la pieza superior de uno de los armazones del resorte de torsión en una gran catapulta lanzapiedras. Las dimensiones eran especificadas por los diseñadores de catapultas en función de las dimensiones de los lados verticales del armazón los cuales, a su vez, quedaban determinados por el diámetro del ovillo de cuerda que constituía el resorte de torsión. No se conoce con exactitud el espesor del tablero citado, pero, probablemente, era igual al diámetro aproximado del ovillo de cuerda. Los constructores de catapultas parecen haber empezado dibujando un rectángulo, uno de cuyos lados fuera igual a la profundidad de la pieza vertical del armazón y el otro de longitud doble de la del anterior. Trazaban, entonces, la diagonal *AG* del rectángulo y, a partir del punto *D*, una paralela *DE* a la diagonal *AG*. Se prolongaba, a continuación, *BG* hasta que cortara a *DE*. Se determinaba el centro del paralelogramo *ADEG* y se trazaba alrededor del mismo un círculo igual al tamaño del ovillo de cuerda (una vez terminada la pieza, el círculo definía un agujero en el que se alojaba el ovillo de cuerda). Se trazaban, entonces, los arcos *DE* y *AG*, cada uno de los cuales tenía un radio tres veces mayor que el diámetro del ovillo de cuerda. Finalmente, los agujeros correspondientes a los espaldones se centraban adecuadamente en los bordes rectos de la pieza en la que llegaban a una profundidad aproximadamente igual a dos tercios de su espesor. De este modo, los técnicos de catapultas habían avanzado sus procedimientos de diseño hasta el punto de incorporar métodos que incorporaban automáticamente las escalas en sus instrucciones para la construcción de catapultas. Una vez seleccionado el emplazamiento y la misión del arma, podía determinarse el tamaño del proyectil y, hecho esto, la fórmula de la catapulta daba el tamaño del ovillo de torsión que se requería. Conocido el diámetro del ovillo, los manuales de construcción, en los que se acumulaban décadas de cuidadosas experiencias, darían, por último, los tamaños de las partes principales de la máquina como múltiplos y submúltiplos del diámetro del ovillo de cuerda. En el momento en el que se construía la catapulta real, partes críticas del armazón de madera se reforzaban con hierro.





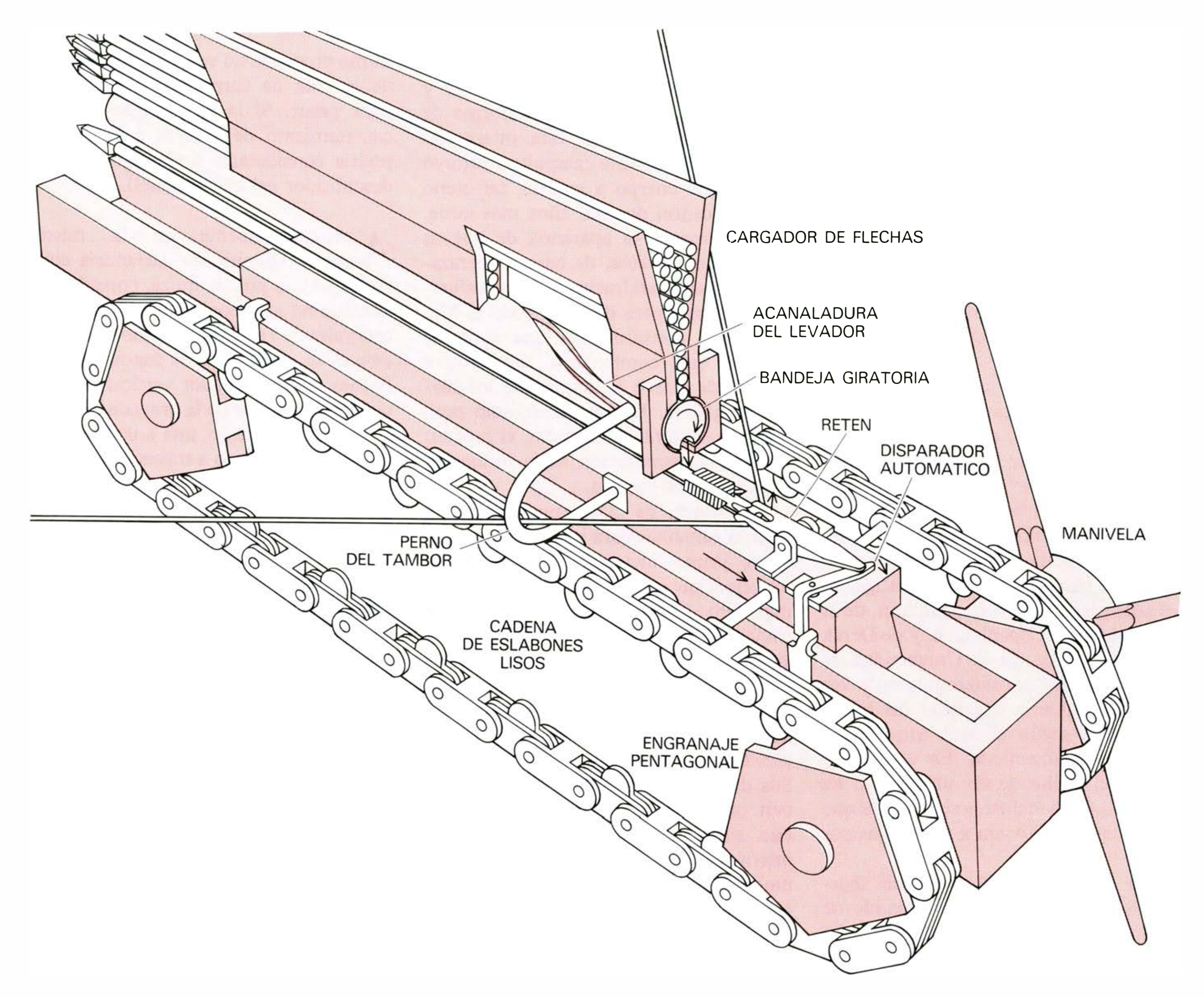
DOS DISEÑOS ALTERNATIVOS. Ctesibio de Alejandría, a mediados del siglo III a. de C., introdujo dos diseños alternativos para catapultas cuya finalidad era sustituir el resorte de torsión habitual. Ambos intentos incorporaban brazos rígidos de arco sujetos a un pivote situado cerca de sus extremos internos. Cuando el arco estaba tensado, los extremos internos de sus brazos se doblaban de tal modo que ejercían presión bien sobre muelles de bronce (*arriba*), bien sobre pistones que se deslizaban en el interior de cilindros herméticamente cerrados (*abajo*). Pero ninguno de los dos diseños proporcionaba una fuerza comparable a la fuerza del arco de torsión y estos intentos se saldaron con un fracaso.

generales dieron ventaja al lado atacante. Hasta los tiempos de la catapulta, los sitiadores se encontraban casi siempre en una situación muy desfavorable. Se supone que la guerra de Troya duró diez años mientras que la lucha entre Atenas y Esparta duró, sin duda, un cuarto de siglo pese a la gran superioridad del ejército espartano. En aquellos días si una ciudad era conquistada, la causa solía ser una traición.

Con la introducción de catapultas y de otras máquinas de guerra que empiezan a aparecer en el mundo occidental, los asedios ganan efectividad. Dionisio de Siracusa primero y, más tarde, Filipo y Alejandro de Macedonia utilizaron torres de asedio elevadas y móviles que se alzaban sobre las murallas de la ciudad y desde las que podía arrojarse fuego que debilitara las almenas mientras enormes arietes, que llegaban a ser empujados por 1800 hombres, golpeaban la parte inferior de la muralla. Evidentemente, las catapultas carecían de una potencia equivalente a la de tales arietes, pero podían derribar muros que no hubieran sido construidos correctamente. Filón de Bizancio, en un manual de artillería escrito hacia el 200 a.C., afirmó que una muralla debía tener una anchura mínima de 4,62 metros para resistir las piedras lanzadas por una catapulta y que era conveniente mantener estas máquinas al menos a 150 metros de distancia mediante fosos y otros obstáculos.

Incluso con buenos muros, las almenas que los coronaban seguían siendo vulnerables. Dado que las almenas no podían ser muy gruesas para no obstaculizar el campo visual, resultaba fácil derribarlas mediante balas de piedra, con lo que los defensores se quedaban sin refugio tras las murallas. Al quedar neutralizado el fuego de estos últimos, los arietes y los equipos de hombres dedicados a socavar los muros podían trabajar con menos impedimentos.

La catapulta tuvo un papel fundamental en los cambios que sufrió la vida urbana en el siglo IV a.C. que la hicieron adquirir un carácter notablemente más precario. Durante sus primeros cinco años de gobierno, Alejandro tomó cinco ciudades importantes y un gran número de ciudades menores. Un pasaje de la *Política* de Aristóteles (el maestro de Alejandro) refleja este cambio. El arquitecto Hippodamus había puesto de moda en Grecia una planificación racional de la ciudad con calles rectas que se entrecruzaban constituyendo bloques cuadrangulares de edificios. Aristóteles objetó que, por lo menos una parte de cada ciudad, debía conservar la tortuosa



LA CATAPULTA MAS COMPLEJA inventada en la Antigüedad fue un arma de repetición diseñada por Dionisio de Alejandría, quien trabajó en el arsenal de Rodas. Tal como muestra este dibujo parcial, las flechas caían, debido a la gravedad, desde un cargador hasta la entalladura de disparo a través de un tambor giratorio que estaba acanalado de tal modo que sólo podía recibir una flecha cada vez. El giro del tambor era controlado por la acanaladura curva de un levador, situada en su superficie, que engarzaba con un dedo de metal montado sobre la corredera. El movimiento de esta última era producido, a su vez, por dos cadenas de eslabones lisos situados a ambos lados de la máquina. Según el texto que conservamos, en el que se describe esta catapulta de repetición, las cadenas se movían sobre prismas de cinco lados, colocados en las dos curvas extremas de las mismas. De acuerdo con la opinión de los autores de este trabajo, parece que estos prismas actuaban como engranajes inversos; dicho de otro modo, la transmisión, sobre la cadena, para la secuencia amartillamiento-disparo se basaba en el engarce entre lengüetas situadas en los eslabones y un engranaje pentagonal que aceptaba las lengüetas. El prisma anterior era movido por una manivela y el gancho de la cuerda del arco se cerraba y abría, en el momento adecuado, mediante clavijas montadas sobre el mango del arma sobre el que se deslizaba la corredera. Por tanto, si se movía la manivela hacia adelante y hacia atrás, el arma podía lanzar flechas automáticamente hasta que el cargador quedaba vacío. (Los dibujos son de Dan Todd.)

estructura de la etapa anterior con el fin de dificultar la entrada de los invasores. Además, escribió, resultaba particularmente importante el diseño de las murallas y el cuidadoso mantenimiento de las mismas en un momento en que "las catapultas y otras máquinas para el asedio de las ciudades habían alcanzado un grado tan elevado de precisión".

El consejo de Aristóteles no era suficiente. La pequeña ciudad estado-democrática estaba muriendo a manos de la nueva tecnología. A medida que sus muros se convertían en vulnerables, iba siendo absorbida por los grandes imperios absolutistas como el de Alejandro. Los combates de carácter democrático habían subrayado el papel de los lanceros, con armas idénticas, que se alineaban hombro contra hombro. Los guerreros tenían idéntico papel y un hombre podía, fácilmente, ocupar el lugar de otro. Esta igualdad en el campo de batalla, así como la igualdad similar de los remeros en los barcos de guerra, constituyó un fuerte apoyo al régimen de la ciudad democrática.

Con la aparición de máquinas militares especializadas se perdió la igualdad de armamento. Se necesitaba una habilidad matemática y técnica especial para construir y mantener una catapulta, y los riesgos que se corrían al manejarla eran inferiores a los de los soldados de infantería. Como resultado de esto empezaron a aparecer jerarquías de especialistas con funciones y prerrogativas particulares. Con el tiempo, a medida que los soldados-ciudadanos empezaban a perder prioridad ante las máquinas, se sintió la necesidad de encontrar una estructura política adecuada a un nivel de desarrollo tecnológico en el que la fuerza muscular humana tenía un papel importante.

El ingeniero antiguo tendió a beneficiarse de esta nueva situación. En la época primitiva no había gozado de un *status* privilegiado. El único de los dioses olímpicos que aparece como un ser

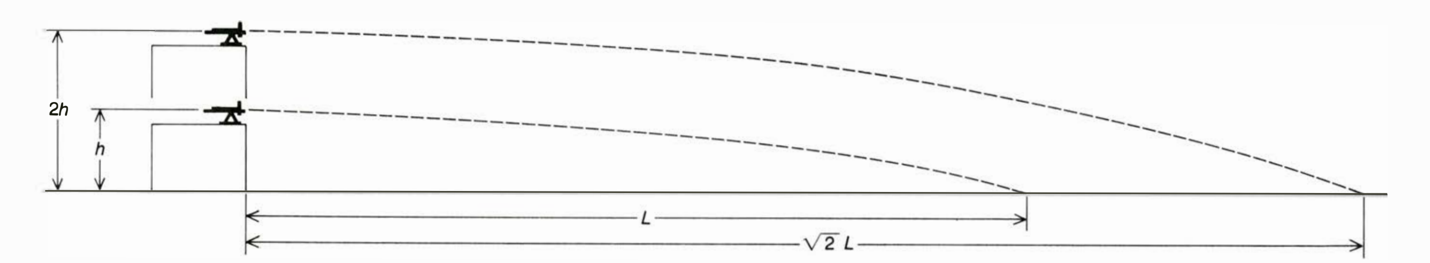
sucio, feo y cojo es Hefesto, que manejaba la forja y realizaba hábiles trabajos de metal (incluso robots). El escultor Fidias, que tuvo el papel de supervisor en la construcción de la Acrópolis bajo Pericles, fue acusado de sacrilegio por atreverse a incorporar su propio retrato a su obra. Más tarde, no obstante, las cosas mejoraron para estos artesanos. No sabemos quiénes fueron los inventores de Dionisio pero conocemos los nombres de tres de los que trabajaron en la corte de Filipo y Alejandro: Polydias, Diades y Karias. En el siglo siguiente surgieron los diseñadores de catapultas y sobreviven los nombres y, al menos, en parte, las obras de Herón, Filón, Bitón, Vitrubio, Ctesibio y otros. La tradición culmina con Arquímedes cuya gran fama, entre sus contemporáneos, se debía sobre todo a sus máquinas de guerra y no a su obra matemática. Las investigaciones recientes de A.G. Drachmann, de la Universidad de Copenhague, y de Derek J. de Solla Price, de la Universidad de Yale, hacen que consideremos cada vez menos probable la exactitud del relato de Plutarco según el cual Arquímedes desdeñaba la ingeniería. En el período romano el hecho de ser un experto en catapultas llegó incluso a servir para que un técnico sobreviviera a las purgas políticas.

Esta elevación en el *status* del ingeniero se debía a la fuerte demanda de catapultas. Estas se convirtieron en parte del utillaje de cualquier fortaleza o equipo de asedio puestos al día, y poco a poco empezaron a ser desplegadas en los combates más móviles del campo de batalla. En el mar pudieron tener un papel en la carrera de armas navales que llevó desde el trirreme, con tres bancos de remos, hasta inmensos barcos que llegaban a tener cuarenta bancos. Evidentemente, la suposición subyacente era que el fuego de la catapulta podía diezmar a la fuerza de abordaje del enemigo cuando su barco se encontraba demasiado lejos para lanzar los garfios o embestir con el espolón. Cuanto mayor era el barco más catapultas podía llevar y más estable resultaba su plataforma de tiro. Por consiguiente, esta interpretación considera que la catapulta sustituyó al combate cuerpo a cuerpo, tal como haría el cañón dos mil años más tarde. Eventualmente, la aparición de nuevas tácticas de combate, de barcos acorazados llamados catafractas y de los esfuerzos de Roma para dominar todo el Mediterráneo constituyeron una serie de factores que, combinados, tendieron a reducir de nuevo el tamaño de los barcos de guerra. En un paralelismo político con el combate en tierra, el proceso implicó la disminución de la influencia del ciudadano-remero.

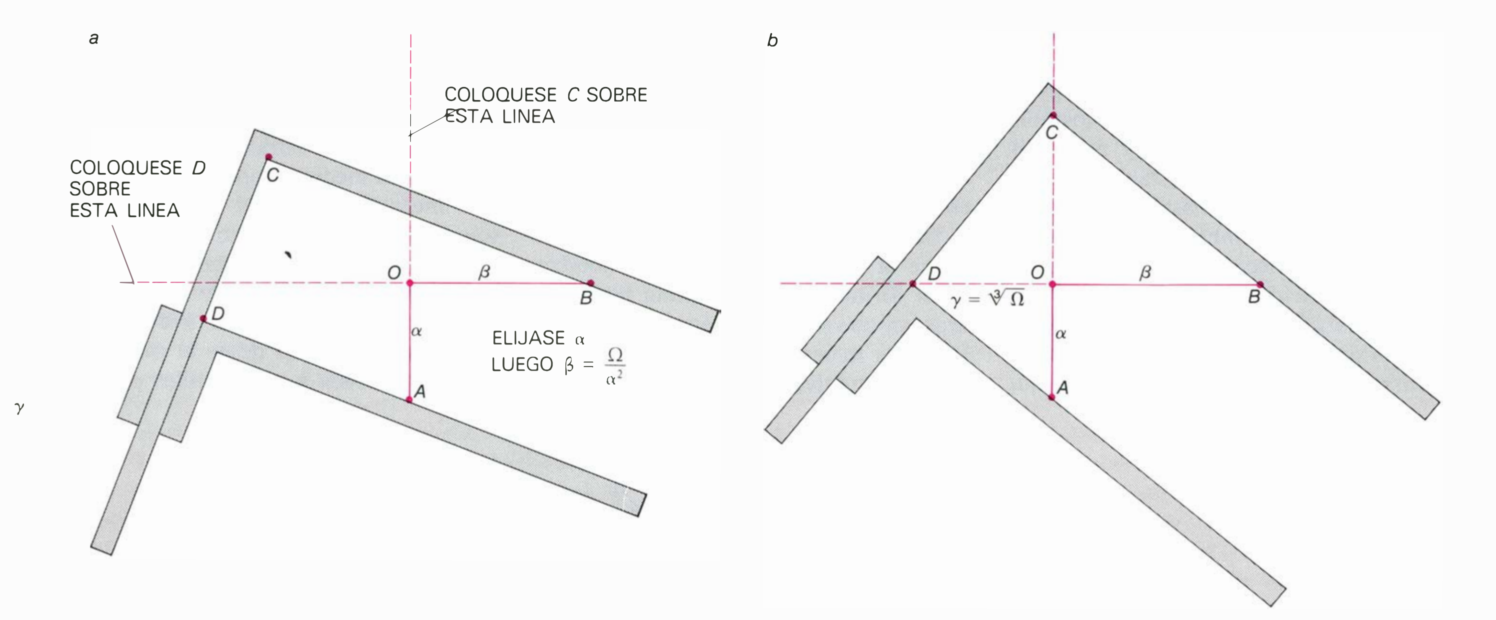
La eficacia de la catapulta movió a realizar esfuerzos para mejorar sus resultados yendo incluso más allá de la introducción del arco a torsión. El ingeniero Ctesibio, por ejemplo, que trabajaba en Alejandría a mediados del siglo III a.C., trató de sustituir las cuerdas fabricadas con cabellos y tendones animales, ya que éstas podían romperse, pudrirse y producir cambios de tensión debidos a la humedad o a la dilatación. Sus dos diseños alternativos incorporaban brazos rígidos, que giraban sobre ejes situados cerca de sus extremos internos y que estaban curvados de tal modo que ejercían una presión, cuando el arco estaba tenso, sobre muelles de martillo de bronce o sobre pistones que se deslizaban en el interior de cilindros herméticamente cerrados [*veánse las ilustraciones de la página 96*] Ni la compresión de los muelles de bronce (que son, desde luego, peores que los de acero a casi todos los efectos) ni la de la pequeña cantidad de aire contenida en los cilindros podían, no obstante, ejercer una fuerza comparable a la del arco de torsión. (En el proceso de investigación Ctesibio descubrió que al producirse la salida del pistón, que había sido introducido en el cilindro a golpes de martillo, salía también "fuego". La llama o el humo procedía, tal vez, de la combustión de la cola de carpintero que utilizaba para pegar. Si la ignición se debía al calentamiento del aire al comprimirse, podría considerarse a Ctesibio como el descubridor del efecto-diesel).

Aproximadamente en la misma época, Dionisio de Alejandría desarrolló lo que puede, quizá, considerarse la máquina más notable de su clase: una catapulta de repetición [*véase la ilustración de la página 97*]. Se depositaban las flechas en un cargador vertical, alimentado por la fuerza de la gravedad, y éstas se trasladaban luego, una a una, a la entalladura de disparo a través de una bandeja giratoria cuyo movimiento era controlado por un sistema constituido por un levador y un tambor de transmisión activado por la corredera. En este sistema, el tambor de transmisión producía un movimiento de vaivén a lo largo del levador, el cual giraba en consecuencia. No se conoce ningún ejemplo más antiguo de levador similar y ninguno tan complejo apareció hasta el siglo XVI. Un solo movimiento de manivela controlaba la bandeja, la corredera, el retén y el gatillo de tal modo que un simple movimiento de la manivela hacia atrás y hacia adelante hacía disparar automáticamente a la máquina hasta que el cargador quedaba vacío. Aquí aparece por primera vez una cadena de eslabones lisos que con frecuencia se ha atribuido a Leonardo da Vinci. Los eslabones de la cadena tenían, probablemente, extensiones que se endentaban con un engranaje invertido; dicho de otro modo, los dientes eran internos, no externos, muy parecidos a los de una sierra de cadena moderna. (Esta interpretación se basa en parte en detalles del texto conservado y en parte en las necesidades mecánicas de la situación.)

La catapulta de repetición no logró sustituir a la catapulta convencional. Su facilidad de manejo y rapidez de disparo se veían contrarrestadas por un alcance



UNO DE LOS EFECTOS del desarrollo de las catapultas fue el forzar a los habitantes de las ciudades a construir murallas más elevadas para que sus máquinas tuvieran mayor alcance. Si se doblaba, por ejemplo, la altura de un muro con ello no se aumentaría el alcance de manera proporcional, pero se vería incrementada la ventaja de los defensores. Este desarrollo en la historia de las fortificaciones sugiere que las catapultas constituían una amenaza mayor que los arietes, ya que las murallas más elevadas resultaban menos estables cuando se las golpeaba por abajo. Las catapultas defensivas se disparaban, probablemente, de un modo más o menos horizontal para obtener una mayor precisión, tal como se muestra en este diagrama. Doblar la altura *(h)* implicaba multiplicar el alcance de la catapulta *(L)* por la raíz cuadrada de 2 (producto indicado en la ilustración por $\sqrt{2}$ *L*).



EXTRACTOR DE RAICES CUBICAS. Se trata de un dispositivo mecánico sencillo, inventado por un geómetra griego anónimo del siglo IV o III a. de C., que hizo posible diseñar a escala una catapulta lanzapiedras gracias a la resolución de una fórmula que establecía que el diámetro del ovillo de cuerda en dáctilos era igual a 1,1 veces la raíz cúbica de 100 veces el peso de la piedra en minas. Para obtener la raíz cúbica de una cantidad Ω, por ejemplo, se empezaba por seleccionar un segmento α y, a continuación, se obtenía el valor β de modo que $\beta = \Omega / \alpha^2$. Se trazaban, acto seguido, los segmentos α y β de modo que el uno fuera perpendicular al otro (*izquierda*). Desplazando la pieza móvil del dispositivo podía llegarse a alinearlo de modo que el punto *C* se encontrara en la prolongación vertical del segmento α y que el punto *D* estuviera en la prolongación horizontal del segmento β (*derecha*). γ, igual a la distancia *OD*, daba, entonces, la solución. Supongamos, por ejemplo, que γ sea igual a la raíz cúbica de 100. Si tomamos α igual a 4, entonces β será igual a 100/16, o sea, 6,25. El extractor de raíces cúbicas, una vez alineado, daría un resultado aproximado para γ de 4,64.

demasiado corto. Por otra parte, su precisión, de modo paradójico, resultó un inconveniente. El artificio concentraba sus disparos en un área tan pequeña a su máximo alcance (unos 200 metros) que no valía la pena abrir fuego a esta distancia incluso contra un grupo pequeño de tropas. (Fue un modelo de estas catapultas de repetición el que logró partir en dos una flecha en la exhibición de tiro realizada por Schramm para el Kaiser.) Los jefes del ejército temían también que su uso supusiera un malgasto de municiones, queja que surgiría de nuevo cuando se inventaran los rifles de repetición dos milenios más tarde.

Otra causa del fracaso de estas interesantes novedades puede encontrarse en la sofisticación de los esfuerzos que los ingenieros estaban aplicando en aquel momento a la catapulta convencional. Su éxito dio lugar a la necesidad de lograr alcances cuya distancia resultara, por lo menos, equivalente a los del enemigo. Se tuvo, por tanto, que ajustar la cantidad de fibra elástica al peso del proyectil. Probablemente, los diseños no se llevaron hasta el límite por el que se conseguía un máximo alcance absoluto, sino sólo hasta el punto en el que se producía un aumento de los costos, menor facilidad de manejo o pérdida de precisión en el disparo debida a factores balísticos.

Una de las etapas cruciales en el diseño del resorte de torsión fue la de establecer la razón existente entre el diámetro y la longitud del ovillo cilíndrico de cuerdas elásticas. Si las cuerdas eran demasiado cortas, desarrollarían una elevada fricción interna y podían no tener suficiente elongación elástica tolerable para no romperse cuando los brazos se tensaban hacia atrás. Si eran demasiado largas, parte de su elasticidad no se utilizaría al estirarse los brazos hasta los límites impuestos por el armazón. Todas las descripciones de catapultas que conservamos implican que se alcanzó una óptima configuración cilíndrica que no se abandonó sino en circunstancias muy especiales: es el caso de las máquinas utilizables sólo para corto alcance que construyó Arquímedes en Siracusa.

Esta optimización del ovillo de cuerda se logró alrededor del 270 a.C., debido quizás a la labor de un grupo de ingenieros griegos que trabajaban para la dinastía ptolemaica en Egipto. Tanto en este país como en Rodas, según Filón, los experimentos de los técnicos en catapultas fueron "fuertemente subvencionados porque en ellos habían monarcas ambiciosos que protegían a los artesanos". Esta fase de las investigaciones terminó con resultados cuantificados que tienen un aire claramente moderno.

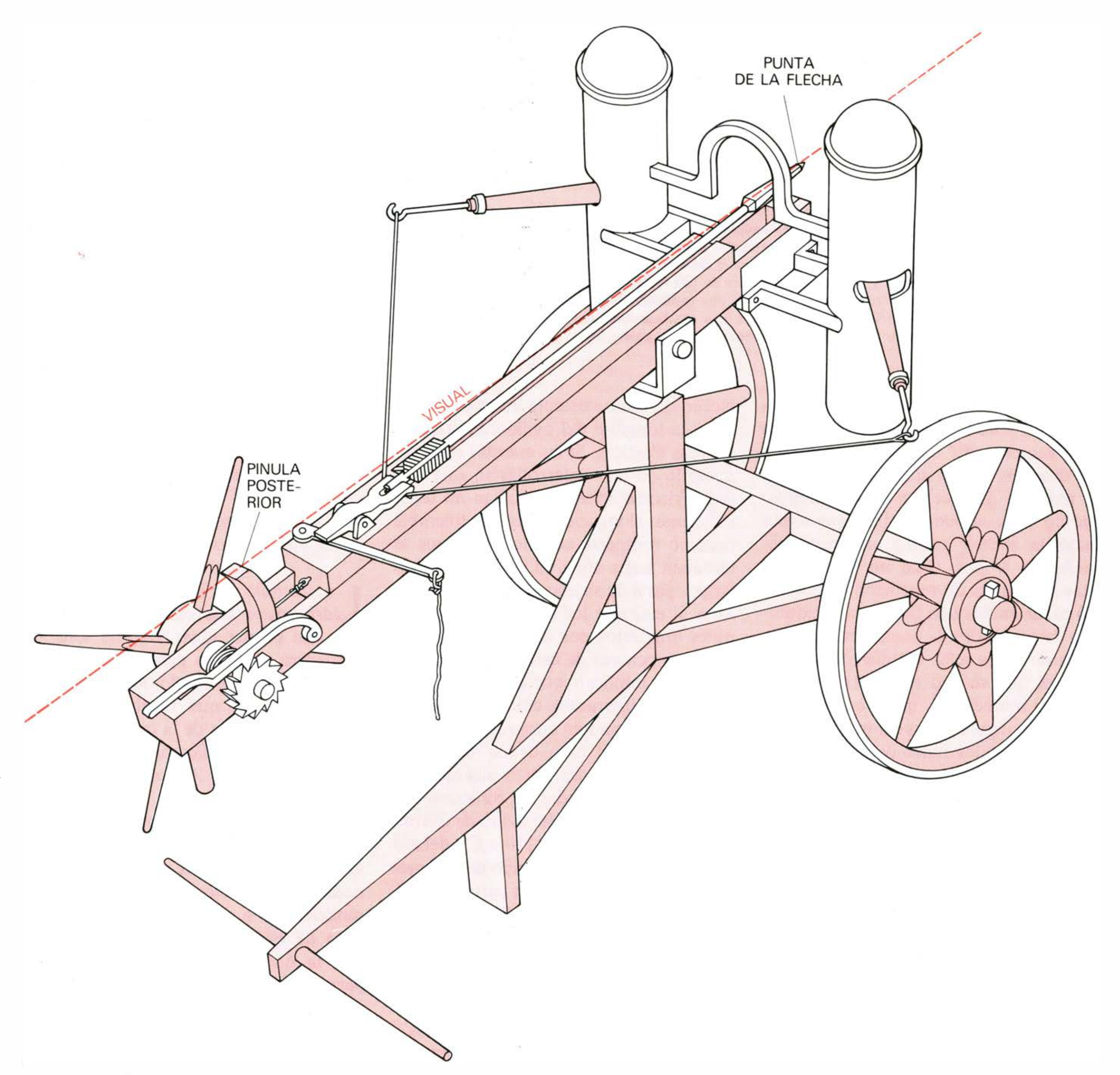
Los resultados se resumieron en dos fórmulas. Para una máquina lanzadora de flechas el diámetro del ovillo de cuerda debe ser, simplemente, 1/9 de la longitud de la flecha. La fórmula, más compleja, de la catapulta lanzapiedras establecía, en términos modernos, que el diámetro del ovillo de cuerda en dáctilos (unos 19.8 milímetros aproximadamente) equivale a 1,1 veces la raíz cúbica de cien veces el peso del proyectil en minas (unos 437 gramos).

La fórmula de la máquina lanzapiedras tiene dos características notables. En primer lugar, da una solución verdadera y precisa para un diseño óptimo. Para comprobarlo, asumamos primeramente (ya que resulta razonable) que los ingenieros de la catapulta deseaban aumentar al máximo el rendimiento de sus máquinas. Por consiguiente, debían también aumentar al máximo la energía cinética de sus proyectiles. Para lograrlo se veían obligados a aumentar al máximo la energía potencial almacenada en los resortes de torsión. La teoría moderna de la elasticidad aplicada al diseño de estos resortes nos informa de que la energía almacenada disponible para el disparo será proporcional a la cantidad de la tensión inicial que se le da al ovillo al enrollarlo a través del armazón, a la tensión adicional causada por la pre-torsión del ovillo, al cuadrado del ángulo que indica la cantidad de torsión adicional lograda al tensar hacia atrás el brazo del arco y al cubo del diámetro del ovillo. El elevar al cubo el diámetro del haz indica que, para expresar el diámetro del resorte de cuerda en función de

la masa del proyectil, es necesario extraer una raíz cúbica.

Téngase en cuenta que, para llegar a este resultado, deben utilizarse los conceptos de energía cinética y potencial, entre los que no llegó a establecerse una relación significativa hasta el siglo XVIII con la obra de Leonhard Euler y Daniel Bernouilli. Se necesita también la teoría de la elasticidad, iniciada por Hooke y Robert Boyle aproximadamente medio siglo antes. Por último, deben utilizarse los principios de la balística que no fueron clarificados hasta la obra de Francesco Cavalieri y Galileo Galilei hacia 1630. Resulta impresionante el que los antiguos ingenieros de catapultas fueran capaces de lograr una fórmula que se mantiene firme a la luz de todos estos desarrollos muy posteriores.

La utilización de una raíz cúbica constituye la segunda característica notable de la fórmula de la catapulta lanzapiedras, ya que se escribió en una época en la que la matemática griega no era aún capaz de resolver de manera total ecuaciones de tercer grado. Hacia el 460 a.C., Hipócrates de Quios (no se trata del famoso médico) afirmó que el volumen de un cubo puede duplicarse con exactitud si pueden encontrarse dos líneas que definan dos medias proporcionales entre dos líneas dadas. En el siglo siguiente sólo se dio un primer paso hacia la solución del problema. Arquitas de Tarento y Eudoxo de Cnido dieron soluciones teóricas elegantes, aunque tridimensionales y embarazosas desde el punto de vista de la física y casi inútiles para el



UN DESARROLLO TARDIO en la historia de la tecnología de la catapulta fue la *cheiroballistra* romana, una máquina relativamente pequeña y provista de ruedas que remonta aproximadamente al siglo I d. C. El armazón de hierro del arma le dio suficiente movilidad para su uso en el campo de batalla y su estructura abierta facilitaba la localización de blancos en movimiento. Probablemente se apuntaba el arma alineando el extremo de la flecha con una pínula, convenientemente elevada, situada en la parte posterior del mango. Las relaciones matemáticas entre la fuerza del arco, el desplazamiento de la corredera y el ángulo de elevación eran tales que, utilizando una visual que pasara por el blanco, la punta de la flecha y la pínula trasera, el tirador podía obtener automáticamente la trayectoria correcta mediante una simple estimación de la distancia, desplazando hacia atrás la corredera, con la ayuda de la manivela, hasta que hubiera girado un número adecuado de dientes del engranaje. El onagro, que lanzaba piedras con un solo brazo, sucedió, a su vez, a la *cheiroballistra*, y constituyó el arma fundamental de la artillería pesada hasta la aparición del trabuquete.

cálculo. Las cosas quedaron así hasta la aparición del arco de torsión.

La mayor parte del grupo siguiente de personajes que se ocuparon de la solución del problema de la raíz cúbica tenía una relación directa o indirecta con las catapultas. Menecmo, según la tradición, fue uno de los maestros de Alejandro Magno y, por consiguiente, estaba presente en el momento y en el lugar en el que el arco de torsión apareció por vez primera. Su solución implica el uso de secciones cónicas que se intersectan, un concepto que parece haber descubierto. Desgraciadamente, no tenemos pruebas de que lo que le llevó a tomar en consideración las cónicas fuera el problema de extraer raíces cúbicas para el diseño de catapultas.

La siguiente solución al problema de la raíz cúbica es la de Eratóstenes, amigo de Arquímedes y natural de Alejandría, ciudad que, entonces, era un centro en la investigación de catapultas. Eratóstenes afirmó explícitamente que la catapulta era la principal razón de orden práctico para ocuparse de los problemas de la raíz cúbica. Podemos suponer que se interesaba por cuestiones de ingeniería ya que Arquímedes le dedicó su libro *Sobre el método*. En esta obra se dan soluciones aproximadas a problemas matemáticos a partir de una aproximación inicial de carácter práctico basada en la ingeniería mecánica. Así, se pesan, por ejemplo, secciones de cuerpos para determinar las razones de sus volúmenes. La solución de Eratóstenes se basaba en un dispositivo mecánico con piezas deslizantes, algo similar al de la ilustración de la página 99.

Los restantes investigadores de la raíz cúbica, entre los que se encuentran Filón de Bizancio, Arquímedes de Siracusa y Herón de Alejandría, fueron famosos por sus trabajos sobre las catapultas. Resulta interesante señalar que la mayor catapulta lanzapiedras documentada, una máquina de tres talentos (78 kilos), fue construida por Arquímedes. Una máquina de este calibre requería la extracción de la raíz cúbica porque no existen raíces naturales en las cantidades requeridas. Arquímedes se vio, también, obligado a abandonar las proporciones normales en las catapultas al construir sus máquinas especiales de corto alcance. La eficacia de éstas constituye un testimonio de su habilidad como ingeniero matemático.

Si se consideran en conjunto, estos estudiosos primitivos del problema de la raíz cúbica se mantienen al margen de la corriente principal en el desarrollo de la matemática griega. En lugar de limitarse al uso de la regla y el compás, idearon dispositivos mecánicos sencillos que les permitieron generar secciones cónicas e incluso curvas de orden más elevado. Instrumental auxiliar, como la simple regla de cálculo que aparece en la ilustración de la página 99, anticipa los compases proporcionales y los sectores de artillero de fines del siglo XVI. La labor de este grupo, no obstante, fue pasada por alto hasta el Renacimiento, momento en el que el desarrollo matemático volvió a ponerse en marcha en el punto aproximado en el que ellos lo habían dejado. *La géométrie* de Descartes, por ejemplo, se inicia con procedimientos y sistemas muy similares a los suyos.

Parece, pues, que los ingenieros de catapultas realizaron experimentos que los arrastraron a un campo en el que los procedimientos matemáticos tradicionales aún no habían penetrado. Resulta bastante fácil, incluso hoy, introducir datos de tercer grado en una curva de segundo grado si los datos son malos o el investigador poco escrupuloso. Se debe, por tanto, sentir una buena dosis de respeto por estos investigadores antiguos. Debieron repetir muchas veces sus experimentos de disparo de catapultas, conservar cuidadosamente sus resultados e interpretar estos últimos con un alto grado de precisión. Los primeros párrafos de la *Belopoieica* de Filón insisten mucho en los procedimientos experimentales y en los resultados obtenidos por los primitivos ingenieros de catapultas: si partimos del punto de vista ventajoso de la moderna teoría de la ingeniería, parece confirmarse plenamente la exactitud de su relato.

Una vez hubieron llegado a un volumen y configuración óptimos del ovillo del resorte de torsión, los técnicos de las catapultas continuaron sus experimentos hasta obtener las dimensiones óptimas para las restantes piezas principales de la máquina. Si los brazos del arco eran demasiado cortos, la fuerza necesaria para montar el arma resultaría excesiva, el trayecto recorrido por la cuerda del arco demasiado breve y, por consiguiente, quedaría limitada su capacidad de transferir energía. Si los brazos del arco eran demasiado largos, retrasarían, debido a su masa incrementada, la acción de los resortes, o bien aumentarían excesivamente el tamaño del arma. Una vez determinada la longitud de los brazos del arco quedaría establecida también, a partir del trayecto recorrido por la cuerda del mismo, la longitud de la corredera y la del mango, y lo mismo puede decirse del resto de la máquina.

De hecho los técnicos de las catapultas escribieron sus textos de tal modo que las dimensiones de sus partes principales se daban como múltiplos del diámetro del resorte. Una vez calculado este diámetro para el tamaño del proyectil deseado, el resto de la máquina quedaba igualmente determinado de acuerdo con una escala adecuada. Los textos conservados que contienen esta información dan testimonio de un nivel de racionalidad tecnológica que no se alcanzó de nuevo hasta el tiempo de la Revolución Industrial.

La última mejora de importancia en el terreno del diseño de catapultas surgió más tarde, en época romana, momento en el que la materia prima del armazón dejó de ser madera para convertirse en hierro. Esta innovación hizo posible una reducción del tamaño, un incremento de los niveles de tensión y una mayor libertad de movimiento de los brazos del arco. Ahora podía montarse la catapulta sobre ruedas y moverse fácilmente de una parte a otra del campo de batalla de acuerdo con la fortuna propicia o adversa del combate. El nuevo armazón abierto facilitaba asimismo la operación de apuntar que se realizaba dentro de estrechos límites con las antiguas estructuras de madera, especialmente cuando los blancos eran móviles y se encontraban muy próximos unos a otros.

Este diseño avanzado de catapulta llegó demasiado tarde para el período de expansión de la civilización romana, pero jugó un papel en la estabilización de las fronteras del Imperio y ayudó a impedir su erosión. No obstante, a medida que avanzaba la decadencia de Roma, las habilidades técnicas necesarias para construir y mantener maquinarias tan complejas parecen haber resultado cada vez menos frecuentes. Se utilizó de un modo creciente una nueva máquina más sencilla, denominada "onagro", dotada sólo de un resorte y de un brazo, que acababa en una cuchara y servía para lanzar piedras. Esta fue el arma fundamental de la artillería pesada de la Edad Media hasta la aparición del trabuquete cuya estructura, más simple, basaba su potencia en la gravedad.

El diseño científico de máquinas complejas con un reajuste experimental deliberado de las dimensiones de sus piezas no apareció de nuevo en la civilización occidental hasta el siglo XVIII. Durante la Edad Antigua los cambios en los que la catapulta tuvo un papel principal anticiparon, de manera sorprendente, resultados que aparecerían de nuevo en las relaciones entre ciencia y tecnología, por una parte, y entre arte militar y sociedad por otra.

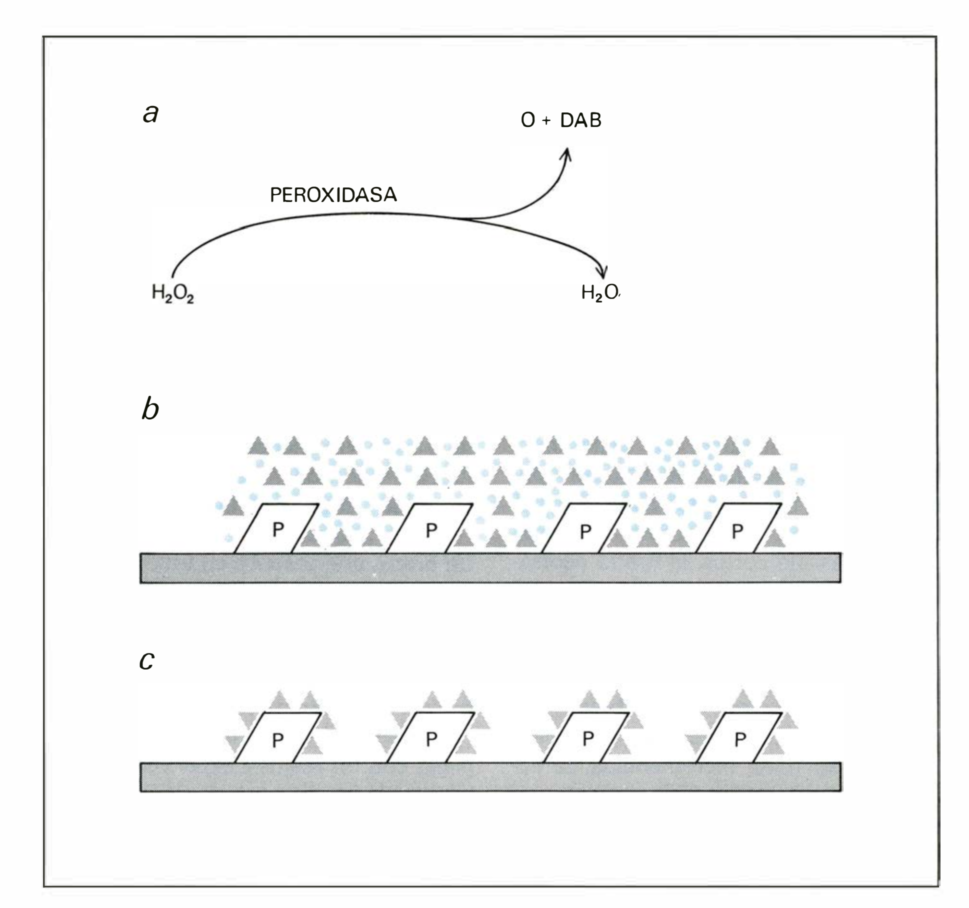
Inmunohistoquímica en microscopía electrónica

El progreso de la biología depende, en gran parte, del avance de nuestro conocimiento de la composición química de aquellas unidades morfológicas que integran el organismo viviente

Antonio Martínez Hernández

El progreso de la biología depende, en buen grado, del avance en nuestro conocimiento de la composición química de aquellas unidades morfológicas que integran la materia viviente. El análisis bioquímico de órganos y tejidos, por sí solo, es insuficiente, pues la homogeneización de un órgano o tejido obscurece su heterogeneidad estructural, heterogeneidad que se extiende hasta un nivel molecular.

Al estudio de la composición de los tejidos y las células en relación con su organización estructural se le denomina histoquímica o citoquímica. La pregunta fundamental que el citoquímico se plantea es la siguiente: dentro del complejo laberinto estructural que es una célula, ¿dónde se encuentra una determinada substancia? Para localizar e identificar dicha substancia se utiliza una reacción específica, generalmente derivada de reacciones que suelen emplearse en química orgánica e inorgánica. Para que una de estas reacciones sea utilizable en histoquímica, únicamente se requiere que el microscopio pueda identificar el producto final de la reacción.



PRINCIPIO DE LA DEMOSTRACION HISTOQUIMICA de la presencia del enzima peroxidasa (*P*). Este enzima cataliza la descomposición del peróxido de hidrógeno (H_2O_2) (agua oxigenada) a agua y oxígeno. Si la solución que contiene el peróxido de hidrógeno contiene también diaminobencidina, el oxígeno liberado por la acción de la peroxidasa oxidará a la diaminobencidina; este óxido de diaminobencidina así formado es insoluble y da un precipitado marrón rojizo, fácilmente identificable (*a*). En *b* y *c* se ilustra la demostración de peroxidasa en una sección montada en un porta. Los puntos de color indican diaminobencidina en solución; los triángulos, precipitado de óxido de diaminobencidina.

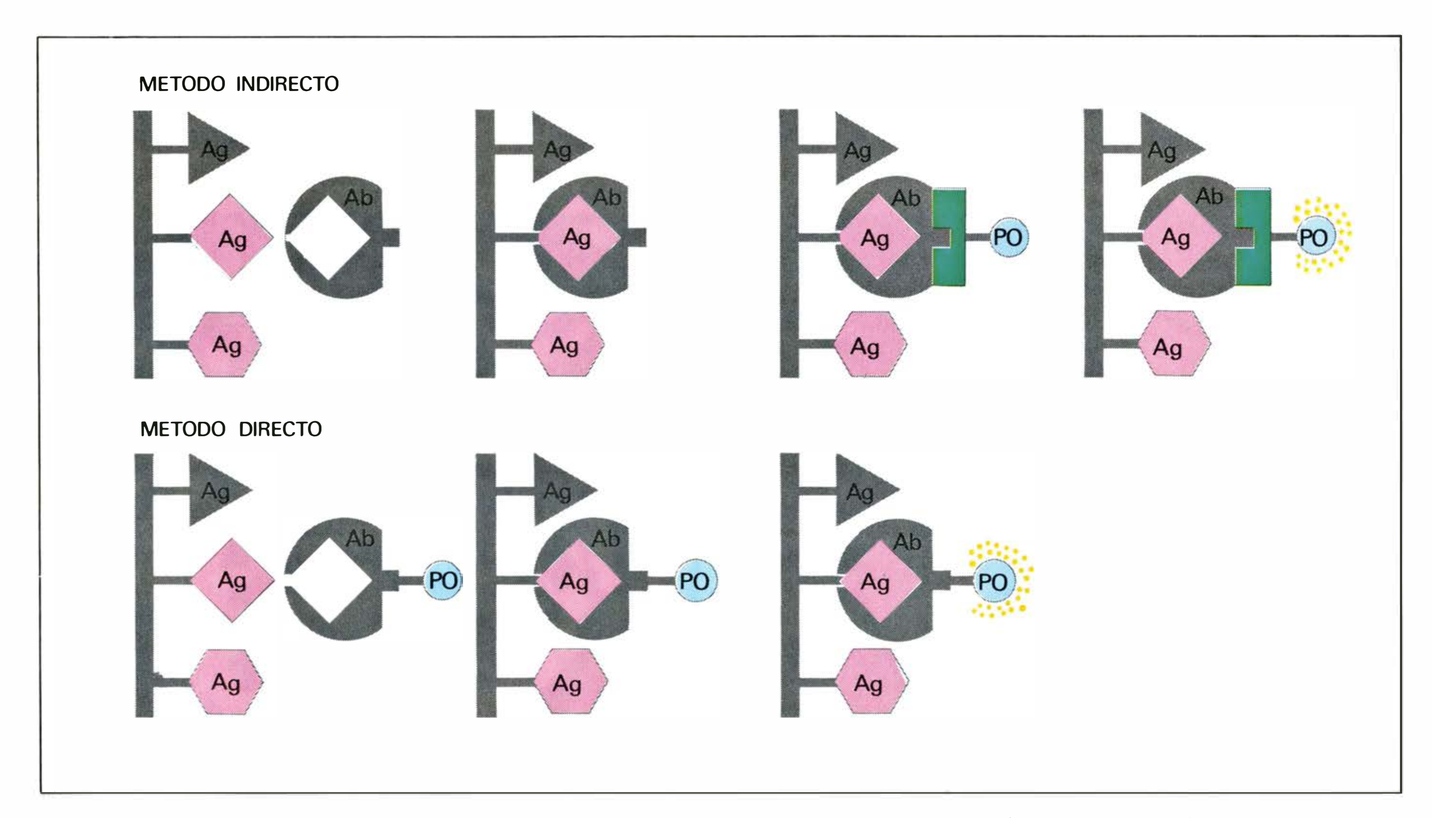
El iniciador de esta modalidad de estudio fue el botánico, farmacéutico y microscopista francés Francois-Vincent Raspail (1794-1878) y es a él a quien corresponde el título de padre de la histoquímica. Raspail, por ejemplo, demostró la presencia de almidón en las semillas, aplicó reacciones usadas en el pasado para demostrar la presencia de proteínas, carbohidratos, etcétera, en órganos y tejidos. Desgraciadamente, pronto decidió abandonar sus fructíferas actividades científicas para aplicar su innegable talento a la política. A pesar de este paternal abandono, la histoquímica siguió progresando, aunque muy lentamente y siempre como una especie de hermana pobre de la histología y de la bioquímica.

Este estado de circunstancias cambió de forma dramática cuando en 1936 el científico francés L. Lison publicó su *Histochemie Animale*, en la cual anunciaba la nueva ciencia de la histoquímica sin destrucción de los tejidos y se proclamaba su profeta. Con la aplicación de los métodos y técnicas descritas en esta verdadera biblia de la histoquímica, la localización de determinadas substancias a nivel tisular o celular se convertía de arte empírico en ciencia experimental. Desde entonces hasta nuestros días, la

histoquímica ha seguido progresando, se ha adaptado a la microscopía electrónica y constituye un instrumento de indiscutible aplicación y utilidad en el estudio de prácticamente cualquier problema biológico.

La histoquímica clásica puede alcanzar gran precisión y especificidad en la identificación de algunos enzimas o fermentos. Los enzimas son substancias que actúan de una forma altamente selectiva sobre una substancia o grupo de substancias (substratos) y facilitan o aceleran una determinada reacción química. La afinidad y especificidad del enzima por su substrato es tan grande que se les ha comparado al acoplamiento de la llave con su cerradura. Pues bien, los histoquímicos han utilizado esta especificidad de los enzimas para definir localizaciones de gran precisión y elegancia. La idea básica estriba en encontrar un substrato (natural o artificial) que, al sufrir la acción específica del enzima, se transforme en un nuevo producto. Si el nuevo producto es insoluble y tiene color, será fácilmente identificable por microscopía óptica. Si es electrónicamente denso ("electron dense") –es decir, posee átomos metálicos– e insoluble, será identificable en microscopía electrónica. Por desgracia, la gran mayoría de substancias cuya localización es de importancia biológica, o carecen de actividad enzimática, o son enzimas que no dan como producto final una substancia insoluble de gran densidad electrónica. Para este tipo de substancias, carentes de acción enzimática, la histoquímica clásica utiliza reacciones que pueden demostrar la presencia de determinados grupos químicos, aldehídos, por ejemplo. En este caso, la reacción histoquímica localiza la presencia de grupos de substancias, pero no nos informa sobre la presencia, o ausencia, de una substancia particular. Sirva de muestra la reacción del ácido periódico Schiff (PAS), que identifica los carbohidratos presentes en las glucoproteínas dando un color rojo magenta. Pero dentro de este grupo de substancias, las glucoproteínas, que comprenden literalmente cientos de proteínas, el PAS no especifica cuál de esas proteínas se halla presente o ausente. El problema que se planteaba para que la histoquímica fuera verdaderamente el instrumento que hiciese posible la conjunción sintética de la histología con la bioquímica, era encontrar métodos de una especificidad absoluta (monoespecificidad) con una sensitividad tal que permitiera determinar la presencia de microgramos o, mejor aún, nanogramos.

En 1941, A. H. Coons y su grupo dieron con una primera solución a este problema. Estos investigadores decidieron aprovechar la increíble especificidad de los anticuerpos (véase "Estructura y funcionamiento de los antígenos histocompatibles", por Bruce A. Cunningham, INVESTIGACION Y CIENCIA, diciembre 1977) para la localización histoquímica. Los anticuerpos son proteínas, gammaglobulinas, sintetizadas por el sistema inmunológico de un animal, cuando se expone este animal a una substancia extraña (antígeno). Por ejemplo, la inyección de un virus a un animal, o a una persona, puede producir una respuesta en el sistema inmunológico de este animal, que eventualmente desemboca en la producción de una o más proteínas (anticuerpos) que reaccionan de forma exclusiva y específica con una determinada parte de las proteínas virales. El ejemplo descrito es un caso particular de respuesta inmunológica,



ESQUEMA DEL METODO DE TINCION utilizando anticuerpos marcados con peroxidasa. En la parte superior se ilustra el método indirecto. La preparación en cuestión reacciona con una solución que contiene el anticuerpo (*Ab*) específico. De entre todos los antígenos (*Ag*) presentes en la preparación (triángulos, cuadrados y hexágonos), el anticuerpo reacciona únicamente con su antígeno específico (cuadrado). El exceso de anticuerpo se elimina por simple lavado en solución tamponada. El tercer paso utiliza gammaglobulina de carnero dirigida contra el anticuerpo de conejo (*Ab, en negro*). Esta gammaglobulina de carnero conjugada con peroxidasa (*PO*) se unirá a su antígeno específico (el primer anticuerpo). En estas condiciones se forma un complejo compuesto por el antígeno que se desea localizar (*cuadrado rojo*) más el anticuerpo de conejo (*negro*) más el anticuerpo de carnero (*en verde*) conjugado con la peroxidasa. Incubando la preparación con este complejo en una solución que contenga peróxido de hidrógeno y diaminobencidina, se producirá un depósito (óxido de diaminobencidina) de color marrón (*amarillo en el dibujo*). (El método directo en la parte inferior, conjuga el primer anticuerpo con la peroxidasa, ahorrando uno de los pasos del método directo. La histoquímica recurre a la especificidad de los enzimas.

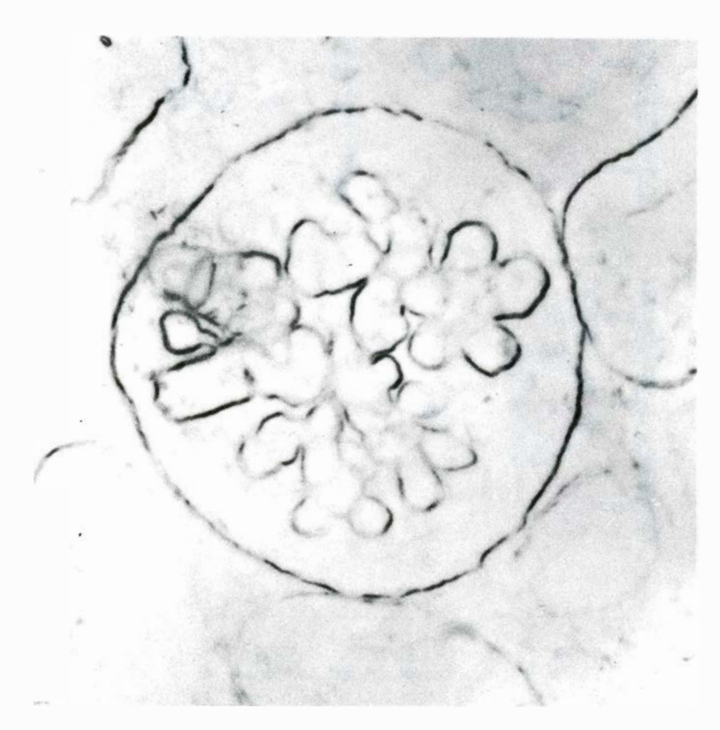
que se ha usado desde hace cientos de años con el nombre de vacunación.

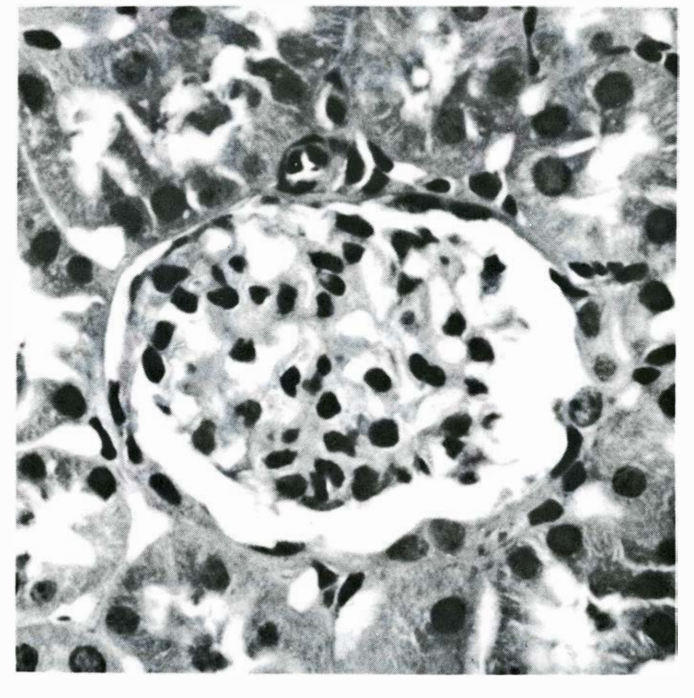
Coons pensó que, si pudiera obtener un anticuerpo específico dirigido contra una substancia determinada, ésta sería la mejor forma de localizar en una sección histológica dicha substancia, y nada más que esa substancia. Por desgracia, los anticuerpos por sí solos no son visibles en el microscopio, pero Coons y su grupo tuvieron la ingeniosa idea de "marcar" los anticuerpos con una substancia, el isocianato de fluoresceína, que aunque por sí misma no es visible, al ser excitada por luz ultravioleta (ondas electromagnéticas cuya longitud de onda no estimula la retina humana y son, por tanto, invisibles) emite una luz verde-amarillenta que es claramente visible. En esencia, pues, la idea de este método es que el anticuerpo, marcado con fluoresceína, reconociera su antígeno de forma específica y formara un complejo antígeno-anticuerpo con él. Todo el anticuerpo que no ha reaccionado con antígeno desaparece con un simple lavado. Cuando un tejido tratado de este modo se examina en un microscopio de luz ultravioleta, la presencia del antígeno específico se revela por la emisión de luz verde-amarillenta característica de la fluoresceína.

Esta técnica esencialmente simple y elegante abrió un campo completamente nuevo en histoquímica. Buena prueba de su utilidad es que, en los últimos treinta años, se ha convertido en técnica imprescindible tanto para el virólogo como para el bioquímico, para el investigador puro como para el mismo patólogo que trata de diagnosticar una enfermedad renal en un paciente.

Con el advenimiento de la microscopía electrónica y sus elevados aumentos y resoluciones, los conocimientos de biología celular dieron un salto gigantesco; los distintos compartimentos celulares dejaron de ser elucubraciones de algunos histólogos para convertirse en una realidad palpable e innegable. El retículo endoplásmico, el complejo de Golgi, las mitocondrias y los ribosomas se revelan como la maquinaria celular; inmediatamente surge la necesidad de definir las funciones de estos orgánulos intracelulares. ¿Qué mejor forma de estudiar la función de uno de estos orgánulos que demostrando cuáles son las substancias que están presentes en los distintos compartimentos, y cómo puede variar el contenido de una substancia de acuerdo con distintos estímulos? Por desgracia, al ser la fluoresceína electrónicamente lúcida, la inmunofluorescencia no puede utilizarse para microscopía electrónica. Por tanto, se impone la búsqueda de métodos que permitan marcar los anticuerpos con substancias densas desde el punto de vista electrónico.

Una primera solución a la inmunohistoquímica para la microscopía electrónica fue la proporcionada por Singer en 1959. Este investigador decidió conjugar los anticuerpos con ferritina en vez de fluoresceína. La ferritina es una proteína de alto peso molecular (650.000 dalton) que contiene cuatro átomos de hierro. Este metal, electrónicamente denso, se halla concentrado en el centro de la molécula de ferritina ocupando los vértices de un cuadrado imaginario. Utilizando esta técnica, queda demostrada la presencia de un determinado antígeno, no por la emisión de luz visible (inmunofluorescencia), sino por la identificación de cuatro puntos electrónicamente opacos (los cuatro átomos de hierro). La idea era excelente y buena prueba de ello es que todavía se sigue utilizando hoy en día. Sin embargo, el método tiene importantes limitaciones y presenta no pocos problemas; por ejemplo, la ferritina tiene una gran tendencia a unirse de forma no específica a otras proteínas y especialmente a los plásticos usados en microscopía electrónica. La consecuencia es que la técnica tiene un gran número de resultados falsos positivos. El gran tamaño de la molécula conjugada, 810.000 dalton (650.000 de la ferritina más 160.000 del anticuerpo), hacen que este conjugado tenga una difusión mínima y no pueda penetrar en células intactas. De esta suerte, su utili-





RIÑON DE RATA. En la microfotografía de la izquierda, aparece el glomérulo renal, en una preparación teñida con hematoxilina y eosina. A la derecha, el mismo glomérulo en una preparación teñida con anticuerpos dirigidos contra la membrana basal glomerular. La reacción positiva identifica a esta membrana basal como una delicada línea festoneada. El círculo central representa la membrana basal de la cápsula de Bowman que rodea al glomérulo. La membrana basal que rodea algunos túbulos adyacentes aparece delineada. La fijación se ha realizado en paraformaldehído-lisina-periodato (método desarrollado por McLean y Nakane para tejidos y células). La inclusión, en metacrilato método indirecto. Aumento: 400 diámetros.

dad queda limitada a substancias extracelulares y a suspensiones de fragmentos celulares. En conjunto, pues, los anticuerpos marcados con ferritina poseen unas ventajas muy menguadas.

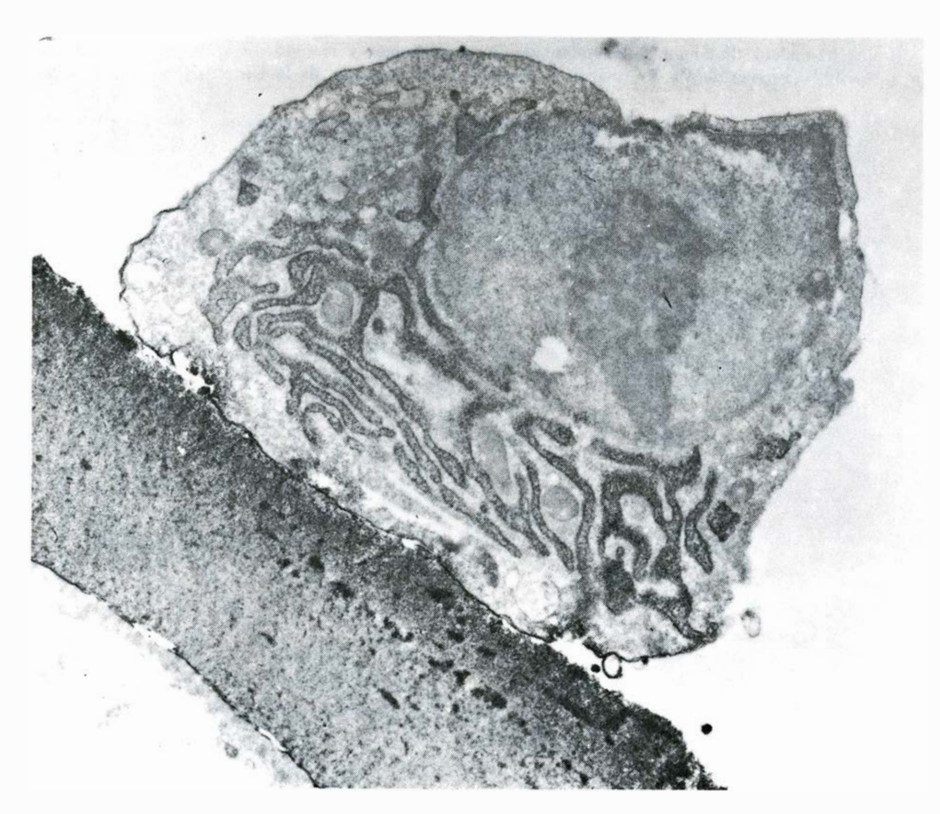
P. K. Nakane y G. B. Pierce aportaron una mejor solución al problema de la localización ultraestructural. En 1966 idearon conjugar los anticuerpos con un enzima y demostrar luego la presencia del enzima usando un substrato adecuado. Después de probar varios enzimas, eligieron una peroxidasa de origen vegetal.

El marcaje de los anticuerpos con peroxidasa tiene múltiples ventajas sobre la ferritina. La peroxidasa carece de la tendencia de la ferritina a unirse de forma no específica con otras substancias; consecuentemente el número de resultados falsos positivos es bajo. El peso molecular de la peroxidasa es de 45.000 dalton, catorce veces menor que el de la ferritina; en consecuencia, su penetración es muy superior. Cada molécula de enzima puede catalizar la conversión de un vasto número de moléculas substrato; así pues, hay un efecto de magnificación enorme, que no se da en el método con ferritina. Por ejemplo, una molécula de antígeno fijará una molécula de anticuerpo que se halla conjugada a una molécula de peroxidasa. La peroxidasa da un producto, el óxido de diaminobencidina, que puede verse tanto en microscopía óptica como en microscopía electrónica, mientras que la ferritina solamente puede ser localizada por microscopía electrónica. Como puede verse, las ventajas del enzima sobre la ferritina son múltiples y significativas. A pesar de estas múltiples ventajas de los anticuerpos marcados con enzimas, la inmunohistoquímica no se halla exenta de problemas y limitaciones. Pasemos revista a algunos de ellos.

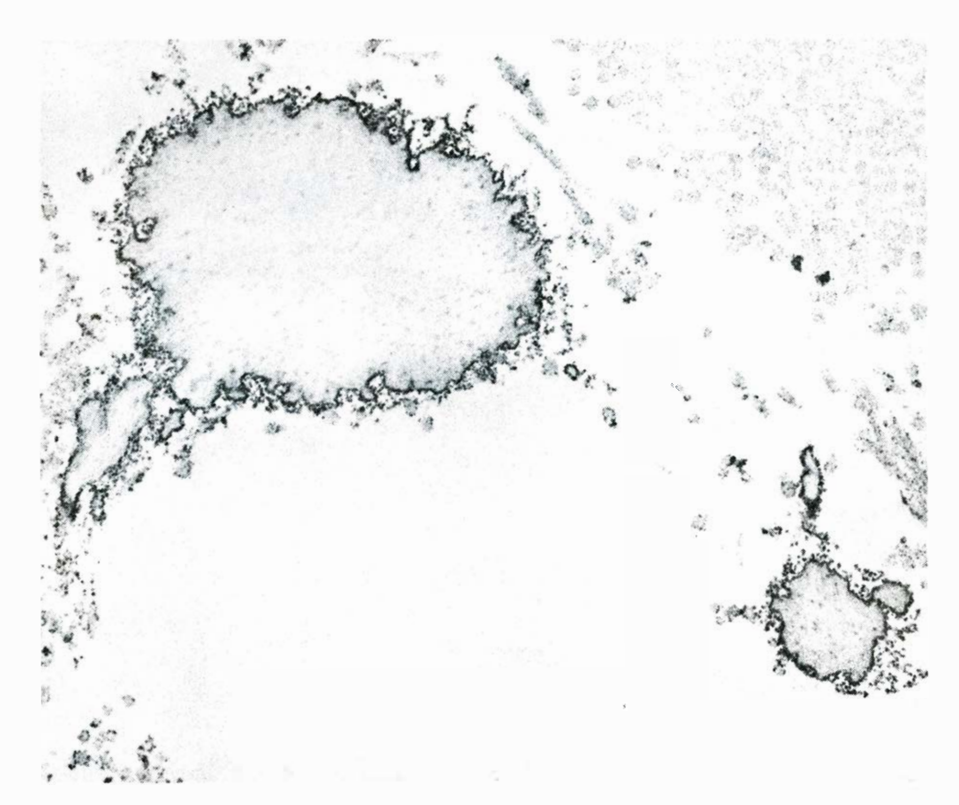
Uno de estos problemas es la fijación de los tejidos. Todas las células contienen una serie de enzimas proteolíticos que pueden autodigerir la célula y destruir toda estructura. A este proceso se le denomina autolisis. Los histólogos han evitado este problema mediante el uso de fijadores. Estos fijadores son substancias capaces de desnaturalizar e inactivar los enzimas intracelulares y, consecuentemente, evitar esta autolisis. Cuanto más completa y rápida sea la desnaturalización de estos enzimas mejor será la apariencia de células y tejidos.

Para obtener una localización inmunohistoquímica adecuada, la fijación que se emplee deberá estabilizar las estructuras celulares y evitar la extracción,

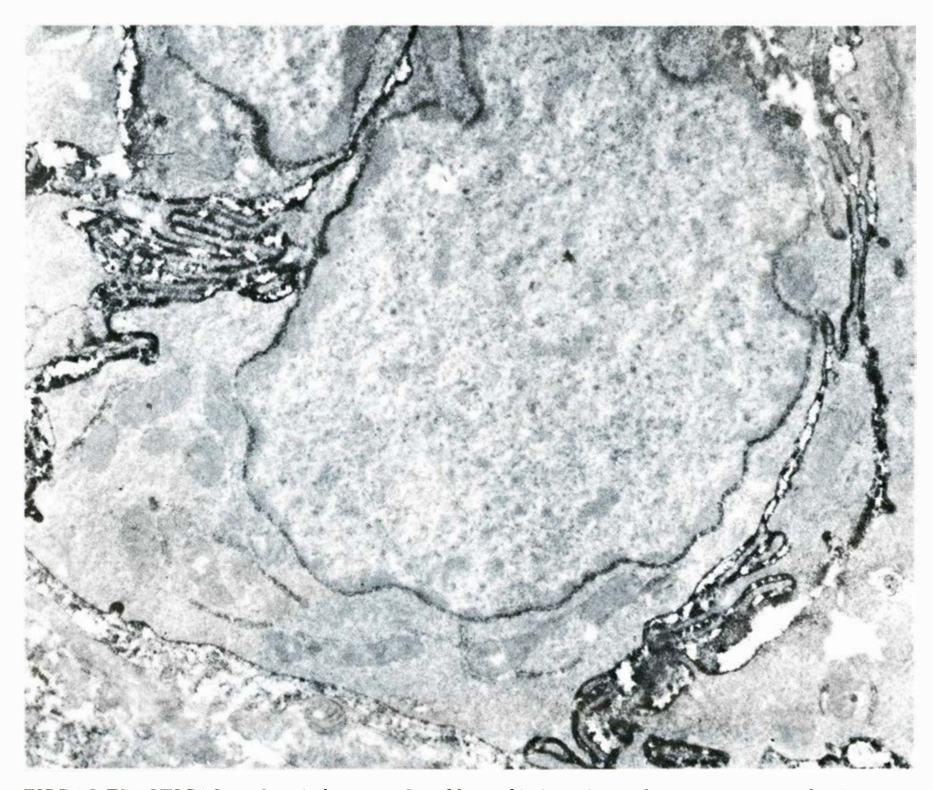




DETALLE DEL SACO VITELINO parietal del ratón, en microfotografías sacadas con microscopio electrónico. En la ilustración superior, las células del saco vitelino parietal descansan sobre una densa membrana basal (membrana de Reichert) que las separa de la placenta. La fijación se ha hecho con glutaraldehído-osmio; la tinción, con acetato de uranio y citrato de plomo. Se ha aumentado 11.000 veces. En la fotografía inferior aparece un detalle del saco vitelino parietal teñido con anticuerpos antimembrana basal. El antígeno queda demostrado por la presencia de un precipitado granular y electrónicamente denso (*gris obscuro en la fotografía*). Este precipitado se encuentra de forma exclusiva en el retículo endoplásmico. El complejo de Golgi (vacuolas en la parte superior e izquierda de la célula) es negativo. La membrana basal es positiva. Estos estudios demuestran que la membrana basal que separa células epiteliales del mesénquima circundante es de origen epitelial, no mesenquimal. La fijación se realizó en paraformaldehido-lisina-periodato. Método directo con Fab'. Aumento x 11.000.



LOCALIZACION DE LA PIEZA SECRETORA de la gammaglobulina *A* en las células epiteliales columnares del yeyuno humano. La pieza secretora se localiza en el espacio perinuclear, retículo endoplasmático y en la superficie de la membrana celular. Estos estudios demuestran que la síntesis de la pieza secretora asociada con la gammaglobulina *A* se realiza en las células epiteliales columnares, especialmente en las que se hallan localizadas en las glándulas criptales. Los métodos de laboratorio seguidos han sido: fijación paraformaldehido-lisina-periodato y método directo con Fab'. Aumento x 16.000. (Cortesía de W. R. Brown, Y. Isobe y P. K. Nakane, de la Universidad de Colorado.)



FIBRAS ELASTICAS en dermis humana. Las fibras elásticas tienen dos componentes: elastina, proteína que constituye el núcleo central amorfo de la fibra elástica, y una serie de fibrillas que rodean a este núcleo. El papel de las microfibrillas en la síntesis, organización y función de las fibras elásticas es un área de gran interés en la fisiología del tejido conjuntivo. La elaboración de los anticuerpos específicos contra las fibrillas periféricas nos ha permitido demostrar que no existe ninguna relación antigénica entre las fibrillas y la elastina. Estos anticuerpos se utilizarán ahora para determinar qué células sintetizan las microfibrillas, así como los mecanismos conducentes a la formación de la fibra elástica. La fijación se ha realizado en paraformaldehído al 4 por ciento, con técnica indirecta; 25.000 aumentos.

desplazamiento e instalación secundaria de los antígenos. Por otra parte, la fijacion no debería desnaturalizar los antígenos, toda vez que el anticuerpo podría no reconocer su antígeno si éste se halla desnaturalizado. Estos dos requisitos constituyen casi una contradicción en los términos, pues un buen fijador puede desnaturalizar cualquier antígeno. Por consiguiente, los fijadores utilizados en inmunohistoquímica en el momento actual representan un compromiso entre una fijación óptima con pérdida total de la antigenicidad, y una antigenicidad intacta con una ultraestructura paupérrima.

Los fijadores más comúnmente usados son aldehídos de cadena corta, tales como el formaldehído y el glutaraldehído. Estos aldehídos reaccionan con proteínas, principalmente con los grupos ε- y α- amino de los aminoácidos que las integran. Por tanto, no es sorprendente que la fijación con estos compuestos tenga habitualmente efectos dañinos sobre la antigenicidad de las proteínas. Ni resulta tampoco extraño que en el momento actual haya un gran interés por el estudio y perfeccionamiento de nuevos fijadores para la inmunohistoquímica. Un primer paso en este sentido ha sido el desarrollo por McLean y Nakane de un fijador (paraformaldehído-lisina-periodato) especialmente diseñado para fijar los tejidos y células; no incide sobre las proteínas (donde radica la antigenicidad), sino sobre los carbohidratos (generalmente no antigénicos). Aunque la enorme utilidad de este fijador resulta obvia, se necesitan todavía grandes mejoras en este sentido.

Por lo que respecta a los anticuerpos, el primer requisito es que deben ser monoespecíficos (dirigidos contra un solo antígeno). Ello implica que el antígeno usado para inmunizar animales sea químicamente puro. Pero cualquiera que sea la pureza del antígeno, no es raro encontrar en el suero de animales hiperinmunizados contaminantes, no detectables por los métodos habituales, que han producido una respuesta, y así, el suero obtenido es en realidad polivalente. Es evidente pues que la gran mayoría de los sueros obtenibles tienen que ser purificados por distintos procedimientos de inmunoadsorción, cromatografía de afinidad y otros para convertirlos en monovalentes.

Otro requisito es la conjugación o enlace covalente entre el anticuerpo y el enzima. La introducción artificial de este nuevo enlace covalente se consigue generalmente mediante una reacción

con una tercera substancia: carbodiimida, dinitrobencenos y glutaraldehído se han usado con cierta frecuencia. Estos reactivos tienen diferente capacidad de conjugación, y con todos ellos se corre el riesgo de desnaturalización y consecuente inactivación de los anticuerpos. Recientemente, Kaowi y Nakane han introducido un nuevo método de conjugación basado en la formación de bases Schiff en los carbohidratos presentes en la peroxidasa, de gran efectividad, que produce un mínimo de desnaturalización de los anticuerpos.

Cualquiera que sea el método de conjugación, la etapa final es la separación de tres moléculas: anticuerpos libres que no se han conjugado, enzima libre y el conjugado enzima-anticuerpo. Una de las formas más simples y eficaces de obtener esta separación es por medio de cromatografía en columna. El conjugado enzima-anticuerpo así conseguido puede emplearse entonces para la localización inmunohistoquímica. En todos los casos se exige una serie de controles, que permitan reconocer artefactos que pueden concluir en resultados falsos. Así, si se emplea el método indirecto, mientras que algunos tejidos se tratarán con suero de conejo inmunizado con el antígeno que se desea localizar, otros tejidos se tratarán con el suero de conejo normal, no inmunizado.

Como puede apreciarse, incluso a partir de esta esquemática revisión, la inmunohistoquímica ultraestructural es una técnica de bastante complejidad, que aún presenta algunos problemas técnicos sin resolver y que ciertamente tan sólo puede seguirse en laboratorios muy especializados. No obstante, la información que puede obtenerse es de gran importancia; ello determina que cada día haya un mayor número de laboratorios que la usen. Algunos ejemplos específicos nos ayudarán a ilustrar lo que hemos dicho.

La membrana basal es una estructura que separa las células epiteliales, tales como las que constituyen la piel, la mucosa que recubre al esófago, estómago, cuello uterino, etcétera, de los tejidos conjuntivos, que actúan como armazón o esqueleto de los distintos órganos. Esta membrana basal es una glucoproteína, que da una reacción positiva con la tinción del PAS. Aunque la función de la membrana basal no esté completamente establecida, hay acuerdo en que fundamentalmente estriba en actuar como un filtro altamente selectivo. Esta función de filtración es muy evidente en el glomérulo renal, una estructura cuya función es filtrar el plasma sanguíneo de aquellos metabolitos indeseables y con ellos formar la orina, y de este modo purificar el plasma. Pues bien, la membrana basal presente en los glomérulos renales sufre cambios radicales, en términos de estructura, en una serie de enfermedades tales como las glomerulonefritis, diabetes, etcétera. Estos cambios en la morfología de la membrana basal se reflejan en cambios en la función de filtrado, resultando en la eliminación de substancias necesarias en la orina, mientras que metabolitos innecesarios o dañinos quedan retenidos en la sangre. Esta situación pronto se hace incompatible con la vida. No es de extrañar pues que tanto los patólogos como los fisiólogos y bioquímicos hayan prestado una considerable atención al estudio de la membrana basal.

Una pregunta de gran importancia para poder entender la fisiopatología de la membrana basal es: ¿qué células sintetizan y secretan la membrana basal, la célula epitelial o la célula mesenquimal? El uso de anticuerpos específicos contra la membrana basal nos permitió demostrar que este antígeno se puede localizar exclusivamente en la célula epitelial adyacente a la membrana basal. Aún más, dentro de estas células el antígeno se halla confinado al retículo endoplasmático, un orgánulo celular cuya función es la síntesis de proteínas. De este y otros estudios similares en varios órganos, la conclusión obtenida es que las células epiteliales son las responsables de la síntesis de la membrana basal sobre la que descansan y tiene que ser por tanto el insulto a la célula epitelial renal, por ejemplo, el responsable de ciertas enfermedades, como el síndrome nefrótico entre otras.

Aspecto sumamente importante de la biología de la membrana basal es su relación con los procesos cancerosos. Puesto que las células epiteliales se hallan separadas de otros tejidos por la membrana basal, y el cáncer es un proceso neoplástico de las células epiteliales, está claro que las células cancerosas tienen que poder atravesar la membrana basal antes de invadir y destruir otros tejidos. Tan es así, que cuando las células malignas permanecen confinadas en el lado epitelial de la membrana basal, se las denomina carcinoma *in situ*; éste es un proceso prácticamente benigno, compatible con una supervivencia indefinida. Sin embargo, cuando las células cancerosas logran destruir la membrana basal, a través de los defectos creados, penetran en el tejido conjuntivo y se convierten en un carcinoma invasivo;



LOCALIZACION del enzima sintetasa de glutamina en el cerebro de rata (corpus striatum). Este enzima, de suma importancia en la detoxificación de la urea, se encuentra exclusivamente en astrocitos. Ninguna otra célula, neuronal o glial, posee este enzima. Esta exclusiva localización proporciona la primera prueba morfológica de la compartimentización de la actividad metabólica cerebral. La fijación se ha realizado en perfusión con formaldehído al 4 por ciento. Método indirecto. Magnificación x 9000. (Reproducción por cortesía del doctor M. Norenberg y Ms. P. Bell, de la Universidad de Colorado.)

este tumor tiene ahora la capacidad de matar al paciente en cuestión de meses o pocos años, a lo sumo. A pesar de que tal relación de la membrana basal con la invasión cancerosa se conoce desde hace más de un siglo, hasta hace tres años no existía prácticamente ningún estudio sobre el catabolismo de la membrana basal. Quizás una de las razones que explique este abandono sea el hecho de que este catabolismo es, tal vez, extraordinariamente lento. Por esto, nuestros estudios se dirigieron a la mama en involución; en este órgano, al interrumpir la lactación, las células epiteliales que segregan la leche mueren y desaparecen en un plazo de cinco a siete días.

Usando anticuerpos contra la membrana basal, hemos podido demostrar que la membrana basal que rodea a las unidades secretoras de la mama desaparece completamente en siete días, al final de cada lactación. Más aún, que la destrucción de la membrana basal no está mediada ni por células inflamatorias ni por fagocitosis, sino que parece ser la consecuencia de la digestión de esta membrana por enzimas proteolíticos segregados por las células epiteliales mamarias. La analogía pues con la lisis de la membrana basal por las células neoplásticas es obvia. Estos estudios todavía en progreso están conduciendo a la purificación del enzima responsable de la lisis de la membrana basal. La purificación de este enzima podría ser de gran importancia en la comprensión y posible prevención de la invasión de tejidos por las células cancerosas.

La inmunohistoquímica está aportando una extraordinaria contribución en los campos más diversos de la medicina y la biología. Así, en el aparato digestivo, la localización de la pieza secretoria de la gammaglobulina *A* en las células epiteliales. En el tejido conectivo, la demostración de dos distintos antígenos en las fibras elásticas.

En el sistema nervioso central, la demostración de que la sintetasa de glutamina se encuentra localizada exclusivamente en los astrocitos ha establecido de forma definitiva el concepto de la compartimentización de las funciones metabólicas cerebrales.

Cualquiera que sea la importancia adscrita a estas contribuciones, no hace falta ser un oráculo para predecir sin miedo a errar que las aportaciones más importantes de la inmunohistoquímica están por venir. Esta esperanza de futuras cosechas es lo que hace fascinante el desarrollo de esta técnica.

Juegos matemáticos

De cómo perturbar el pasado, retrasar el futuro y de otras indebidas manipulaciones del tiempo

Martin Gardner

"Es imposible meditar sobre el tiempo y sobre el misterio de la transición creativa de la naturaleza sin sentirse agobiado por las abrumadoras limitaciones de la inteligencia humana."

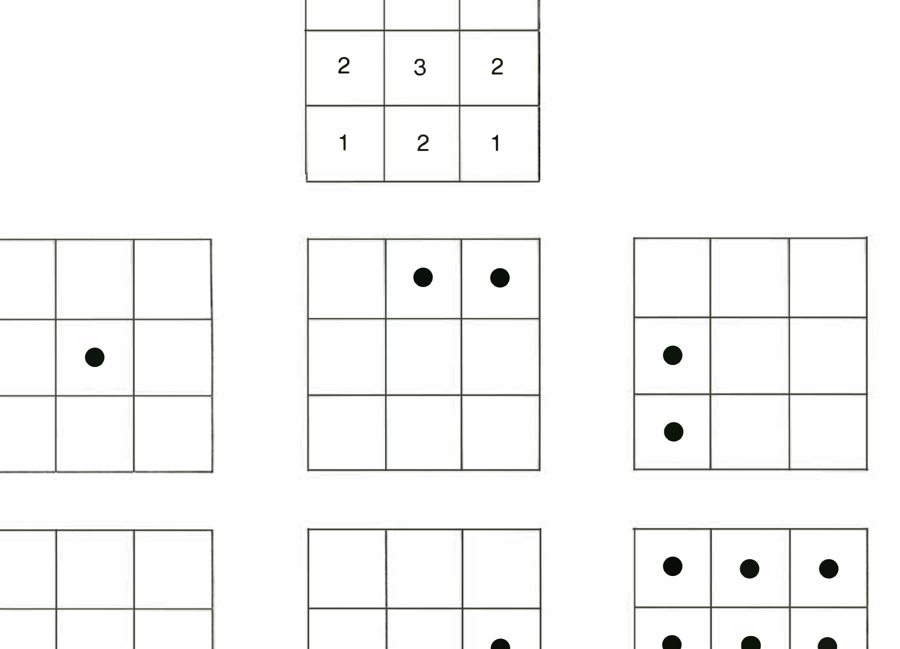
—Alfred North Whitehead, *The Concept of Nature*

Los físicos de los últimos años han manifestado gran interés por elucidar si en el dominio de las partículas elementales existen o no sucesos donde no sea posible invertir el sentido del tiempo; es decir, sucesos para los que imaginar una inversión del sentido de movimiento de todas las partículas implicadas en ellos equivaldría a imaginar un suceso de imposible aparición en la naturaleza. Richard Feynman ha sugerido un enfoque de la mecánica cuántica según el cual las antipartículas se contemplan como partículas que momentáneamente viajan hacia atrás en el tiempo. Hay cosmólogos que especulan sobre la posibilidad de que existan dos universos, ligados de forma que todos los sucesos que se manifiestan en uno de ellos aparecen en el otro en orden inverso respecto al transcurso del tiempo. En cada universo los organismos inteligentes vivirían normalmente, desde el pasado hacia el futuro, pero si los organismos de cada universo pudiesen observar sucesos pertenecientes al otro (lo que muchos físicos consideran imposible) descubrirían que la sucesión de tales eventos se desarrolla en orden contrario al del universo propio. Incluso se ha conjeturado que si nuestro universo dejase de expandirse y comenzase a contraerse, se produciría una inversión en el sentido de progresión del tiempo, aunque, a decir verdad, no está muy claro qué se quiere expresar con ello. La mayoría de las especulaciones a este respecto son sumamente recientes; los lectores podrán ver examinadas muchas de ellas en cuatro capítulos añadidos a mi *Ambidextrous Universe* (cuya traducción española lleva el título *Izquierda y derecha en el Cosmos*; Alianza Editorial, Madrid) que ahora está de nuevo en prensa, en edición revisada y ampliada que publicará Charles Scribner's Sons.

Este mes me detendré a examinar dos extravagantes cuestiones, concernientes ambas a la noción de tiempo, y que no figuran en mi libro. De hecho, tales cuestiones preocupan tan poco a los científicos que tan sólo filósofos y autores de ciencia ficción han tenido algo que decir al respecto: ¿tiene sentido hablar de detención del tiempo? ¿Tiene sentido hablar de alteración del pasado?

Ninguna de estas cuestiones deberá ser confundida con la conocida tesis de la relatividad del tiempo. Newton imaginaba que el universo estaba perfundido de un único tiempo absoluto, que podría simbolizarse mediante un reloj imaginario situado en algún lejano lugar del espacio (y posiblemente, en el exterior del cosmos). Este hipotético reloj serviría para comparar temporalmente cualesquiera dos sucesos acontecidos en el universo, y para medir el tiempo transcurrido entre ellos. Tal noción de tiempo es bastante adecuada en el seno de un único sistema de referencia inercial, como por ejemplo, en pequeñas regiones de la superficie terrestre, pero falla aparatosamente cuando se manejan varios sistemas inerciales moviéndose relativamente unos respecto a otros a velocidades muy grandes. Según la teoría de la relatividad, si una nave espacial hubiera de viajar desde el nuestro hasta otro sistema solar, desplazándose a velocidades cercanas a la de la luz, los sucesos que acontecieran a bordo de ella se desarrollarían allí mucho más lentamente que en la Tierra. En cierto sentido, podría decirse que la nave viaja a través del tiempo hacia el futuro. Para los tripulantes de la nave, el viaje de ida y vuelta podría parecerles una experiencia de unos pocos años; en cambio, al retornar a la Tierra encontrarían que en ella habrían transcurrido unos cuantos siglos.

La idea de que los cambios producidos en distintos lugares del universo puedan desarrollarse con distinto ritmo en el tiempo es muchísimo más antigua que la teoría de la relatividad. Ya la teología escolástica de la Edad Media admitía que los ángeles eran seres inteligentes e inmateriales que vivían según un tiempo diferente al de las criaturas terrestres, y se consideraba que la propia



Estrategias vencedoras para turnablock de tres por tres

existencia de Dios estaba totalmente desligada de cualquier noción de tiempo. En el primer acto del drama de Byron, *Cain, A Mystery*, el ángel caído, Lucifer, dice:

With us acts are exempt from time,
and we
Can crowd eternity into an hour
Or stretch an hour into eternity:
We breathe not by a mortal
measurement-
But that's a mystery

(En nosotros están los actos desconectados del tiempo
y podemos
apiñar en una hora la eternidad completa
o dilatar una hora hasta la eternidad.
No respiramos en las medidas de los mortales-
Aunque esto es un misterio.)

En el siglo XX, centenares de cuentos de ciencia ficción han jugado con la relatividad del tiempo respecto de sistemas inerciales diferentes; pero la idea de que el tiempo pueda acelerarse o retardarse en distintos lugares de nuestro universo está en el corazón de muchos mitos antiguos. Una leyenda medieval nos habla de un monje que entró en trance durante varios minutos, encantado por los trinos de un pájaro mágico. Cuando el pájaro deja de cantar, el monje descubre que han transcurrido varios siglos. En una leyenda musulmana, Mahoma es transportado al séptimo cielo por una yegua. Tras permanecer largo tiempo en él, el profeta retorna justo a tiempo de sujetar una jarra de agua que su montura había derribado al instante de comenzar su ascensión.

En la literatura norteamericana, el más conocido relato sobre personas que permanecen dormidas durante tiempos fantásticamente largos es "Rip Van Winkle", de Washington Irving, cuyo protagonista comprueba que tras un sueño aparentemente normal han transcurrido realmente dos décadas. Gyneth, la hija del Rey Arturo, durmió durante 500 años a causa de un encantamiento de Merlín. Hay leyendas de durmientes semejantes a estas en todas las literaturas. H. G. Wells se sirvió de este recurso en *When the Sleeper Wakes*, y en las narraciones de ciencia ficción es cosa corriente someter a los astronautas a tratamientos de sueño en hibernación para que puedan sobrevivir a viajes interestelares cuya duración rebasaría la de sus vidas ordinarias. En una narración de Wells, "The New Accelerator", un científico descubre un procedimiento de acelerar el tiempo biológico de un sujeto hasta el extremo de que el mundo parece quedar detenido. También este recurso se ha empleado con frecuencia en ciencia ficción.

Pero la cuestión que estamos examinando aquí no es cómo puede variar el tiempo, sino en qué sentido podría decirse que el tiempo se ha detenido completamente. Como es obvio, no es absurdo afirmar que en cierto lugar del universo ha cesado todo movimiento, tanto si ese lugar existe como si no. En la teoría de la relatividad, la velocidad de la luz es límite superior inalcanzable para todo objeto dotado de masa. Si una nave espacial pudiese alcanzar la velocidad de la luz (cosa que la teoría prohíbe, pues la masa de la nave tendría que hacerse infinita), el tiempo se detendría a bordo de ella, en el sentido de que ya no podría producirse cambio ninguno. Medidos en tiempo terrestre, podrían ser necesarios 100 años para que la nave alcanzase su destino; mas para los astronautas que la tripulan, tal destino se alcanzaría intantáneamente. Se podría imaginar también un trozo de materia —y por qué no, un ser humano— llevado a tan baja temperatura (por medios hoy por hoy desconocidos) que incluso llegaran a detenerse los movimientos de las partículas subatómicas. Para tal trozo de materia podría decirse que el tiempo se ha detenido. Aunque la verdad, resulta difícil imaginar que tal trozo de materia no llegara a aniquilarse en el proceso.

La idea de la detención del tiempo no crea problema ninguno a los autores de fantasías, no constreñidos al mundo real. Por ejemplo, en uno de los cuentos de Frank L. Baum de la serie *American Fairy Tales* (que está ahora disponible en edición facsímil publicada por Dover), el titulado "The Capture of Father Time" nos cuenta como un muchachito atrapa al Tiempo con un lazo, y durante un rato todo movimiento, excepto los del muchacho y los del Padre Tiempo, queda totalmente interrumpido. En el capítulo 22 de la novela de James Branch Cabell *Jurgen: A Comedy of Justice*, el tiempo exterior duerme mientras Jurgen disfruta de una placentera estancia en Cocaigne junto a la Reina Anaitis. Más adelante, Jurgen mira a los ojos del Dios de su abuela y queda transido y completamente inmóvil durante 37 días. En un cuento de Jorge Luis Borges, un escritor va a ser fusilado. Entre la orden de "fuego" y la muerte del escritor, Dios detiene todo el tiempo exterior al cerebro del autor, concediéndole un año para que finalice su obra maestra.

Muchos ejemplos parecidos, tomados de las antiguas leyendas o de la moderna literatura, muestran que la idea de detención del tiempo en algún lugar del universo no es lógicamente incoherente.

1	2	1	4	1	2	1	8
2	3	2	8	2	3	2	12
1	2	1	4	1	2	1	8
4	8	4	6	4	8	4	11
1	2	1	4	1	2	1	8
2	3	2	8	2	3	2	12
1	2	1	4	1	2	1	8
8	12	8	11	8	12	8	13

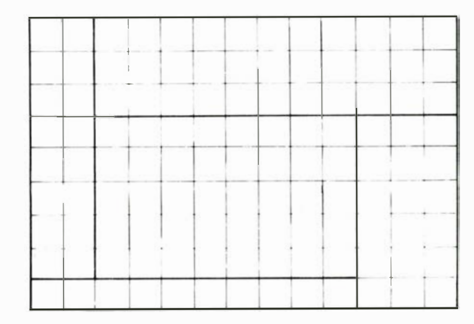
Tabla de estrategias para el turnablock

PERSISTENCIA	NUMERO
1	10
2	25
3	39
4	77
5	679
6	6,788
7	68,889
8	2,677, 889
9	26,888,999
10	3,778,888,999
11	277,777,788,888,899

Mínimos números de persistencia menor o igual que 11

Mas ¿qué decir de la detención del tiempo en todo el universo? ¿Tendría algún sentido concebir que todo movimiento cesara, y tras una pausa, todo recomenzara?

Si se supone que existe un observador exterior, —¿por qué no un dios?— que contempla el universo desde un lugar del hipertiempo, entonces, evidentemente, la noción de detención del tiempo tiene sentido, lo mismo que de existir un dios en el hiperespacio tendría sentido el concepto de volver todo del revés. La historia de nuestro universo podría ser como una película tridimensional que un dios proyecta para su entretenimiento. Cuando el dios apaga el proyector y se va a hacer otras cosas, todo se detiene, y es posible que transcurran unos cuantos milenios antes de que regrese y lo ponga otra vez en marcha. (Después de todo, ¿qué son unos cuantos milenios para un dios?) Por lo que se nos alcanza, podrían muy bien haber transcurrido unos cuantos miles de millones de siglos de hipertiempo en la pausa mecanográfica entre la primera y segunda palabra de esta frase.



Rectángulo de mínima área: 117

Supongamos, sin embargo, prohibido todo observador exterior, y que el término "universo" se define como "todo cuanto es". ¿Cabe todavía la posibilidad de atribuirle significado a la idea de que el tiempo haga una pausa? Aunque la mayoría de los científicos y filósofos dirían que no, unos cuantos han tomado partida por la afirmativa. Por ejemplo, en "Time without Change" (*The Journal of Philosophy*, vol. 66, n.º 12, pags. 363-381; 19 de junio de 1969) Sydney S. Shoemaker, que actualmente es profesor de filosofía en la Universidad de Cornell, presenta un extraordinario razonamiento en apoyo de la posibilidad de detención del tiempo.

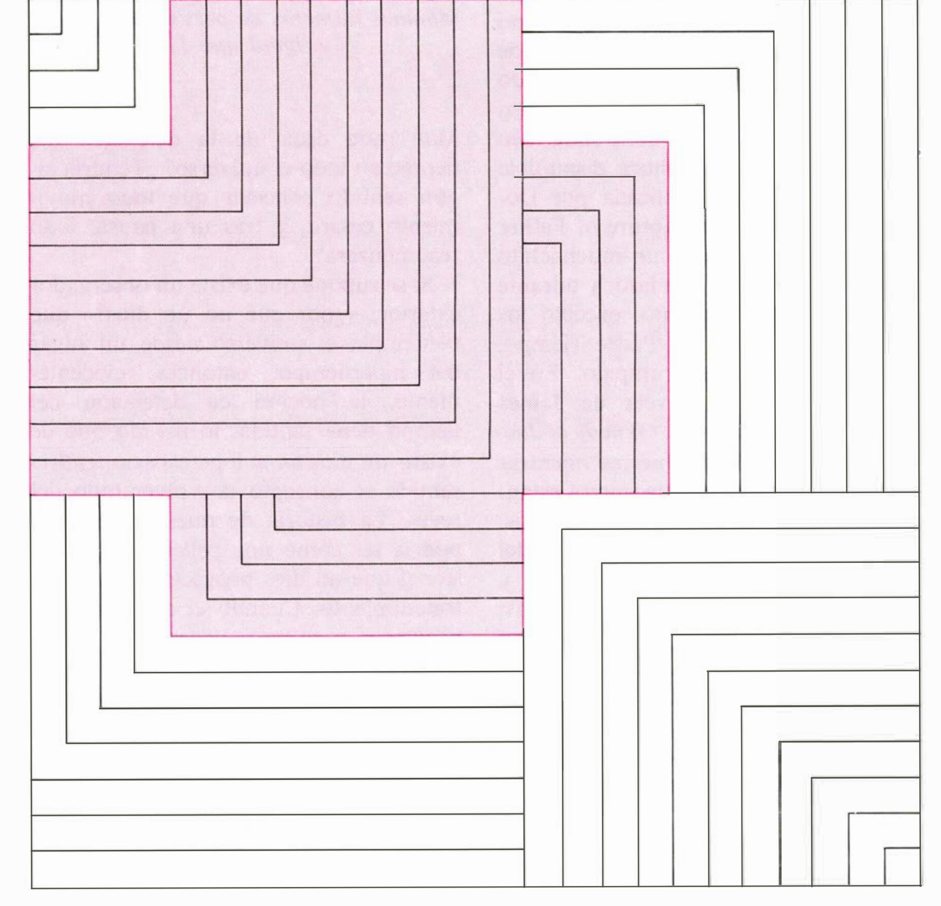
Shoemaker no se refiere al mundo real, sino a posibles mundos inventados por él para demostrar que se puede atribuir significado razonable a la idea de detención universal del tiempo, y propone para ello varios mundos de este tipo, basados todos ellos en la misma idea. Describiré aquí solamente uno de ellos, en forma ligeramente novelesca.

Imaginemos un universo dividido en regiones *A*, *B* y *C*. En tiempos normales, los habitantes de cada región pueden observar a los habitantes de las otras dos, y comunicarse con ellos. Sin embargo, de vez en cuando, un misterioso halo púrpura impregna una de las regiones. El halo dura invariablemente una semana, y está siempre seguido por un año en el cual cesa todo cambio en la región afectada. Dicho de otra forma, durante un año no sucede absolutamente nada en esa región. Shoemaker llama congelación local a este fenómeno. Cuando la región retorna a la vida, sus habitantes están completamente ignorantes de que haya transcurrido tiempo ninguno, pero llegan a saber por sus vecinos que, de acuerdo con los relojes de las otras dos regiones, ha pasado todo un año. A los habitantes de la región que ha sufrido la congelación local les parece que en las otras dos se han producido cambios instantáneos. Como dice Shoemaker: "La gente y los objetos parecerán haberse movido de forma discontinua o haberse esfumado de repente, o por el contrario, haberse materializado de la nada; los retoños se habrán transformado bruscamente en árboles crecidos, y así sucesivamente".

La historia de las tres regiones consigna que las congelaciones locales se han producido ya miles de veces, y han venido precedidas siempre por el halo púrpura, cuya duración siempre ha sido de una semana. Supongamos ahora que por primera vez en la historia se presente simultáneamente la luz púrpura en las tres regiones *A*, *B* y *C*, durante una semana. ¿No sería razonable –pregunta Shoemaker– que los científicos de las tres regiones llegaran a la conclusión de que el tiempo se ha detenido durante un año (signifique esto lo que signifique) en todo el universo, a pesar de que ninguna mente inteligente haya podido percatarse de ello?

Shoemaker examina después diversas objeciones que podría suscitar su tesis, refutándolas ingeniosamente. Los lectores curiosos pueden consultar su artículo y leer después un análisis técnico de la cuestión en el quinto capítulo de la obra de G. Schlesinger *Confirmation and Confirmability* (Oxford University Press, 1974). Schlesinger coincide en afirmar con Shoemaker que a la sentencia "Ha transcurrido un periodo de tiempo *t* durante el cual no ha sucedido absolutamente nada" puede asignársele un significado empírica y lógicamente coherente. Obsérvese también que razonamientos semejantes al anterior aplicados a otros posibles mundos pueden proporcionar significado a frases como "todas las cosas del universo han sido vueltas boca abajo, o se han duplicado, o se han convertido en sus imágenes especulares, etc. etc".

Si ya resulta extraña la detención del tiempo, todavía lo es más la idea de modificar el pasado. Es frecuente ver especulaciones literarias sobre lo que hubiera podido ocurrir si el pasado hubiese tomado un giro diferente. De esta cuestión se ocupa un drama de J. B. Priestley, *Dangerous Corner*; los cuentos de ciencia-ficción y de la literatura en general que tratan del "qué hubiera ocurrido si" son innumerables. En todas las historias de "máquinas del tiempo", cada vez que los viajeros llegan al pasado, éste es perturbado. La única manera de resolver las contradicciones lógicas originadas en semejante premisa es postular un universo que se bifurca en distintas ramas cada vez que se penetra en el pasado. Dicho de otra forma, mientras el tiempo prosigue gorgeando en la vieja rama (la frase es de Emily Dickinson) el tiempo crea nuevos trinos en la nueva cantando un futuro diferente. Sin em-



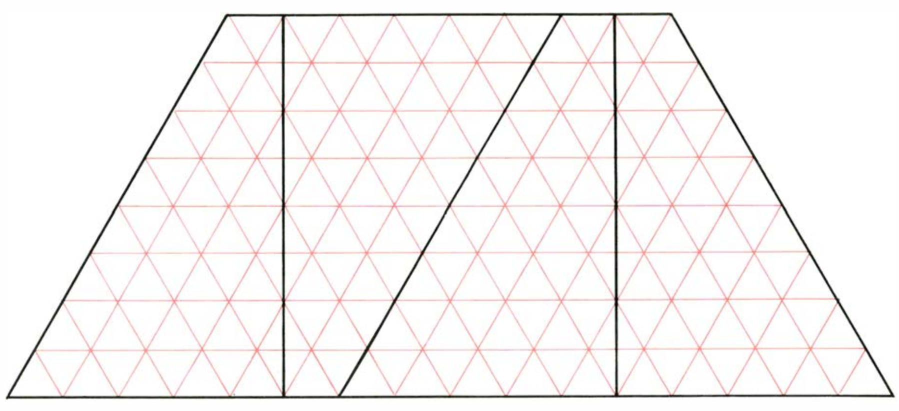
Solución de un problema del mes pasado

bargo, cuando hablo de perturbar el pasado me estoy refiriendo a perturbarlo en todo un único universo, sin ramificaciones ni universos paralelos. (Las pseudoalteraciones del pasado, como la continua manipulación y adaptación de la historia que tan mordazmente satiriza George Orwell en su *1984*, no vienen al caso.) Dado este contexto, ¿puede un suceso, una vez ocurrido, ser forzado a no haber sucedido?

La cuestión es incluso anterior a Aristóteles, quien en su *Etica* (Libro 6) escribe: "Es preciso observar que nada ya pasado puede ser objeto de elección; por ejemplo, nadie puede elegir haber saqueado Troya; pues nadie *delibera* sobre el pasado, sino al contrario, sobre lo que es futuro y susceptible de llegar a ser, ya que lo pasado no es capaz de no haber tenido lugar. Por ello Agathon tiene razón cuando dice: "Pues sólo de esto carece incluso Dios, obligar a deshacerse a las cosas que ya han sido efectuadas".

Santo Tomás de Aquino creía que Dios estaba desligado del tiempo y era, por tanto, capaz de ver el pasado y todo el futuro de su creación en un cegador instante. (Incluso aunque los seres humanos tienen albedrío genuino, Dios sabe cómo elegirá cada uno; de esta forma buscaba el de Aquino armonizar predestinación y libre albedrío.) Según Santo Tomás, a Dios no le es posible realizar cosas absolutamente imposibles, a saber, las cosas que conllevan contradicciones lógicas. Por ejemplo, Dios no podría formar una criatura que fuera a un tiempo caballo y ser humano (es decir, un ser humano completo y un caballo completo; no se habla de combinaciones de partes de ambos, como los míticos centauros) pues ello conllevaría la contradicción de suponer que una creatura puede ser al mismo tiempo racional e irracional. Análogamente, Dios no podría alterar el pasado. Tal equivaldría a afirmar que el saqueo de Troya ocurrió y no ocurrió. Santo Tomás coincidía con Aristóteles en afirmar que el pasado ha de ser por siempre lo que ha sido, y este punto de vista llegó a ser doctrina oficial del Escolasticismo medieval. No se trata tanto de que la omnipotencia de Dios quede limitada por la ley de no-contradicción, sino más propiamente, de que tal ley forma parte de la naturaleza de Dios. "Es preferible decir, escribió Santo Tomás, que lo que implica contradicciones no puede ser realizado, en lugar, de decir que Dios no puede hacerlo."

Edwyn Bevan, al examinar la noción de tiempo en su libro *Symbolism and Belief* encuentra extraño que Santo Tomás negase a Dios la capacidad de modificar el pasado y al mismo tiempo le concibiese capaz de alterar el futuro. En la décima cuestión de la *Summa Theologica* (Ia. 10, artículo 5.3), escribió el Aquinate: "En efecto, como el ser y el no ser del angel dependen de la omnipotencia divina, puede Dios en absoluto hacer que el angel no exista en lo futuro, pero no puede hacer que no exista mientras existe, o que no haya existido después de que existió." Pues al haber sugerido Santo Tomás que el pasado es inalterable, pero que para Dios el futuro no lo es, argumenta Bevan, lo que está haciendo es colocar a Dios en alguna categoría temporal, contradiciendo su afirmación de que Dios es extratemporal.


Un triamante descompuesto en cuatro regiones idénticas

No conozco ningún científico ni filósofo secular que haya defendido seriamente que es posible alterar el pasado, pero sí ha habido una pequeña minoría de teólogos que han mantenido esta posibilidad. El más importante de todos ellos fue Pedro Damiano, celoso reformador italiano de la Iglesia Católica, del siglo XI. En *Sobre la omnipotencia divina*, su más controvertido tratado, Damiano argumentaba que Dios no está en modo alguno constreñido por la ley de no-contradicción, y que su poder le consiente hacer todas las cosas, aún las contradictorias, como por ejemplo, modificar el pasado. Aunque Damiano, que en sus comienzos fue monje ermitaño, argumentaba con la máxima habilidad sus extremadas opiniones, consideraba superfluo todo razonamiento, cuya única utilidad era la de apoyar la teología revelada. Parece como si él –lo mismo que la Reina Blanca de Lewis Carroll– estuviera dispuesto a defender el derecho de todos a creer en seis cosas imposibles antes del desayuno. (Damiano fue también uno de los grandes promotores de la autoflagelación como forma de penitencia, práctica que llegó a hacer furor durante su vida; hasta el punto de que algunos monjes llegaron a producirse la muerte a latigazos.)

No conozco en la literatura fantástica nada que pueda ilustrar mejor la creencia de Damiano en la posibilidad de alterar el pasado que un cuento de Lord Dunsany, uno de mis favoritos. Se titula "The King That Was Not", y puede verse en un temprano libro de Dunsany, colección de cuentos prodigiosos, titulado *Time and the Gods*. Comienza dicho cuento como sigue: "El país de Runazar no tenía Rey ni lo tuvo jamás; y es la ley del país de Runazar que, en vista de que nunca tuvo Rey, jamás tendría uno. Por consiguiente, en Runazar son los sacerdotes quienes dominan, y quienes enseñan al pueblo que nunca en Runazar hubo Rey".

El comienzo del segundo párrafo es sorprendente: "Althazar, Rey de Runazar..." El cuento prosigue, narrando como Althazar ordenó a los escultores de su reino que cincelasen en mármol estatuas de los dioses. Su orden fue obedecida, pero al descubrir las grandes estatuas, resultó que sus rostros se asemejaban mucho al del rey. Althazar quedó complacido, y recompensó a sus escultores espléndidamente con oro; en cambio, en lo alto del monte Pegana (el Olimpo, en los cuentos de Dunsany) los dioses montaron en cólera. Uno de ellos, Mung, se inclinó hacia adelante para hacer su signo contra Althazar, pero los otros dioses lo detuvieron: "No lo mates, pues no basta con que Althazar muera. Quien ha hecho los rostros de los dioses como los rostros de los mortales, ni siquiera ha podido existir".

"¿Estamos hablando de Althazar, uno de los Reyes?", preguntó un dios.

"No, jamás hemos hablado de él."

"¿Hemos soñado nosotros con un tal Althazar?"

"Nunca. Jamás hemos soñado con él".

Al pie de Pegana, en el palacio real, Althazar se desvaneció súbitamente de la memoria de los dioses, y de esta forma "se convirtió en una nada que ya no era una cosa ni que jamás lo fue". Cuando los sacerdotes y cortesanos entraron en el salón del trono, solamente se encontraron un manto y una corona.

"Los dioses han arrojado de sí un trozo de vestidura", dijo uno de los sacerdotes, "y ¡mirad! de los dedos de los dioses se ha escurrido un pequeño anillo."

He aquí las soluciones a los problemas del mes pasado:

1. Los cinco siguientes movimientos son suficientes para conseguir que los cuatro vasos de los pozos de la mesa giratoria queden todos boca arriba o boca abajo.

(1) Se introducen las manos en cualquier par de pozos diagonalmente opuestos. Si los dos vasos no estuvieran ya boca arriba, ajústelos de manera que lo estén. Si al hacerlo no suena el timbre:

(2) Haga girar la mesa y pruebe esta vez en cualquier par de pozos adyacentes. Si ambos vasos están vueltos hacia arriba déjelos así; en caso contrario, invierta el vaso que está boca abajo. Si todavía no suena el timbre, sabe usted entonces que hay tres vasos vueltos hacia arriba y uno hacia abajo.

(3) Haga girar la mesa y pruebe con cualquier par diagonal de pozos. Si uno de los vasos está boca abajo, inviértalo y sonará el timbre. Si ambos están boca arriba, invierta uno de ellos, con lo que los vasos quedarán dispuestos según el esquema

Arriba	Abajo
Arriba	Abajo

(4) Haga girar la mesa y pruebe en cualquier par de pozos adyacentes. Invierta ambos vasos. Si ambos estaban orientados en la misma dirección, el timbre sonará ahora; de lo contrario, los vasos quedarán en el siguiente esquema

Arriba	Abajo
Abajo	Arriba

(5) Gire la mesa, pruebe en cualquier par de pozos diagonales e invierta ambos vasos. El timbre sonará.

Ronald L. Graham, de los Bell Laboratories, y Persi Diaconis, de la Stanford University, han examinado conjuntamente dos procedimientos para generalizar el problema. Uno de ellos, suponer que el jugador tiene más de dos manos. Cuando el número de vasos sea n mayor que cuatro, el problema es resoluble con n - 2 manos, si y solamente si n no es número primo. Por consiguiente, para cinco vasos (5 es número primo) el problema no es resoluble con 5-2, o sea 3 manos. Puede ocurrir cuando n sea número compuesto que para ciertos valores de n el problema sea resoluble con menos de n - 2 manos. El problema puede generalizarse también usando, en vez de vasos, objetos que puedan adoptar más de dos posiciones. Así pues, el problema de la mesa giratoria conduce a profundas cuestiones de combinatoria, que yo sepa, todavía no estudiadas.

2. La estrategia ganadora para el primer jugador del "turnablock", juego inventado por John Horton Conway, requiere numerar las nueve casillas como se muestran en la ilustración de la página 110. El jugador debe tan sólo efectuar movimientos que dejen fichas negras en aquellas casillas que tengan "suma nim" igual a cero. La suma nim de una colección de números se obtiene ordinariamente escribiendo los números en notación binaria y sumando después sin llevar. Si todos los dígitos del total son ceros, entonces la suma nim es también cero. Conway sugiere un procedimiento más sencillo para hallar la suma nim: expresar cada uno de los números a sumar en forma de suma de potencias de 2, y simplificar después los pares de potencias de igual grado. Si no queda ninguna potencia, entonces la suma de nim es cero; de lo contrario, la suma nim es simplemente la suma de las restantes potencias. Consideremos por ejemplo 1 + 5 + 12. Al escribir cada número de esta expresión en suma de potencias de dos, distintas entre sí, resulta 1 + 4 + 1 + 8 + 4. Los pares de "unos" y "cuatros" se simplifican mutuamente, con lo que resulta una suma nim igual a 8. Análogamente, veamos qué ocurre con 1 + 2 + 5 + 6. Esta expresión se convierte en 1 + 2 + 4 + 1 + 4 + 2. Todas las potencias se simplifican mutuamente, y la suma de nim resulta ser cero.

Los puntos negros de la ilustración de la página 110 indican los bloques que es posible mover en la primera jugada, pertenecientes a estrategias vencedoras. Dicho de otra forma, tras cada uno de tales movimientos quedan fichas negras en casillas cuya suma nim es cero. Cualquiera que sea entonces la jugada del oponente quedarán necesariamente fichas negras situadas en casillas de suma nim no nula. El primer jugador podrá siempre responder con una jugada de suma nim igual a cero, y prosiguiendo de esta forma es seguro que alcanzará la victoria.

La ilustración superior de la parte derecha de la página 111, facilitada por el propio Conway, da la numeración de las casillas de todos los tableros de "turnablock" rectangulares cuyos lados van desde uno hasta ocho. La numeración de un tablero de a por b menor que el tablero de orden ocho viene dado por el rectángulo a por b encajado en la esquina superior izquierda de la matriz. Los números de las casillas se obtienen mediante un proceso llamado "multiplicación nim", que el lector puede ver expuesta en el libro de Conway *On Numbers and Games* (Academic Press, 1976, pág. 52).

3. El mínimo número que en base 10 tiene persistencia multiplicativa igual a 5 es el 679. La tabla de la ilustración inferior de la parte izquierda de la página 111 da los números mínimos de persistencia multiplicativa igual a 11 o menor. La tabla se ha tomado de un artículo de N.J.A. Sloane, "The Persistence of a Number", aparecido en *Journal of Recreational Mathematics* (vol. 6, n.º 2, págs. 97-98; primavera de 1973).

El mínimo número que en base 10 tiene persistencia aditiva igual a 4 es el 19.999.999.999.999.999.999.999. La suma de las cifras de este número es 199, que es el mínimo número de persistencia aditiva igual a 3. La suma de los dígitos de 199 es 19, que a su vez es el mínimo número de persistencia aditiva igual a 2. La suma de las cifras de 19 es 10, mínimo número de persistencia aditiva igual a 1. Con mayor generalidad, la segunda etapa de la sucesión que da el número de mínima persistencia aditiva igual a k, da el número de mínima persistencia aditiva igual a k-1. Todos los números de este tipo empiezan por 1 y están seguidos por "nueves". Por consiguiente, el número mínimo de persistencia aditiva igual a 5 es el formado por la cifra 1 seguida por 2.222.222.222.222.222.222.222 "nueves". Puede verse un análisis más completo del problema de Harvey J. Hinden: "The Additive Persistence of a Number" (*Journal of Recreational Mathematics*, vol. 7, n.º 2, págs. 134-135; primavera de 1974).

4. La ilustración de la superior de la página 112 muestra el rectángulo de mínima área (117) que puede descomponerse en cinco rectángulos más pequeños de lados enteros, de forma que ningún par de lados de ningún par de rectángulos interiores sean idénticos. Tal configuración es la única posible.

5. (1) Se muestra la solución en la figura inferior de la página 112. El rompecabezas es adaptación de una ilusión óptica tomada de Charles H. Paraquin, *Eye Teasers: Optical Illusion Puzzles* (Sterling Publishing Co., Inc. 1978).

(2) la regla para obtener la tercera configuración de cada fila consiste en superponer las dos primeras y suprimir cualesquiera líneas comunes que pudiesen tener. Por consiguiente, la figura a situar en el extremo de la tercera fila es sencillamente un cuadrado.

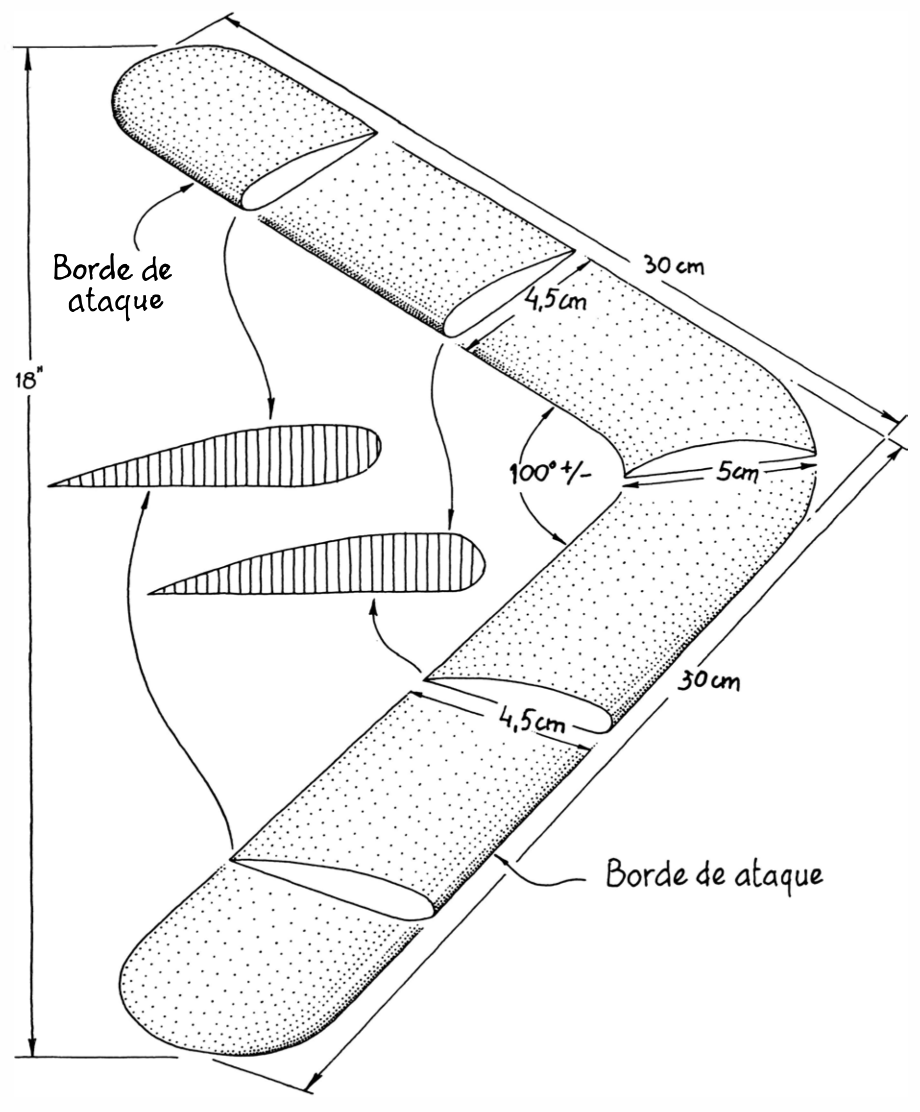
(3) La ilustración de la página 113 muestra la otra única forma que un triamante puede ser descompuesto en cuatro regiones idénticas.

Taller y laboratorio

El encanto de los bumerangs. Cómo construirlos y mantenerlos en vuelo

Jearl Walker

El bumerang es, sin duda, uno de los instrumentos más antiguos de los usados por el hombre, bien como arma ofensiva o bien como juguete. Se cree casi con seguridad que fue descubierto por los indígenas australianos, al parecer de forma accidental. De cualquier manera, también se tienen noticias de su uso por otros muchos pueblos primitivos en diversos puntos de la tierra. Si lanzamos un palo cualquiera al aire, caerá al suelo, y no demasiado lejos de nosotros; en cambio, un bumerang puede recorrer más de 200 metros (en su recorrido de ida y vuelta) e incluso, si lo lanza un experto, puede ir dirigido con tan extremada precisión como para lograr dar a un enemigo o bien hacer blanco, si es que de un juego se trata. El bumerang se inventó probablemente como arma de ataque diseñada para asegurar una trayectoria rectilínea, pero a la mayoría de la gente le pareció más interesante la versión deportiva australiana del bumerang capaz de retornar a su punto de lanzamiento. Por paradójico que resulte, el bumerang de vuelo rectilíneo es aerodinámicamente más complicado que el de vuelo con retorno. Aunque tanto un modelo como otro son muy antiguos, el investigador aficionado encontrará aquí un tema de profundización al objeto de esclarecer las distintas características de vuelo del ingenio.



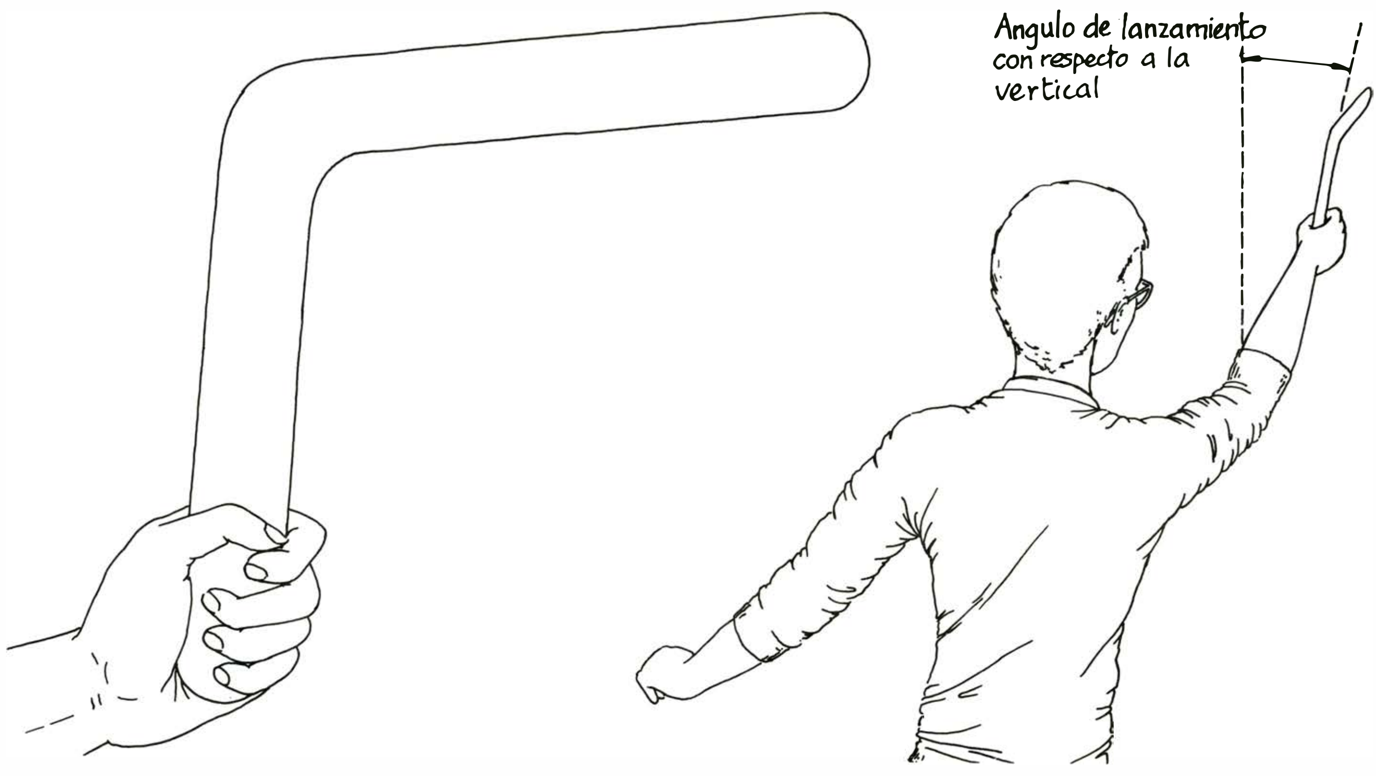
Diseño "Gem" para un bumerang por Herb Smith

Ocasionalmente pueden hallarse buenos ejemplares en tiendas de deportes, pero la mayoría de los bumerangs comerciales están fabricados en serie y no vuelan demasiado bien. Es más, muchos de ellos ni siquiera retornan. Si el lector quiere divertirse con uno, mejor será que empiece por construírselo. Sólo así podrá hacer las variaciones necesarias para determinar qué factores influyen en el vuelo de un bumerang. Un material óptimo para el tronco básico es abedul del Báltico; se aconseja recortar la forma del bumerang de una plancha contrachapada de 6 a 10 milímetros de espesor con cinco o más capas, del tipo de las empleadas en la construcción de embarcaciones. Este tipo de planchas contrachapadas resisten el desgaste por uso y el ataque del agua; son densas, cosa que hace que el bumerang pese más de lo que corresponde a su tamaño.

Recorte en un cartón o cartulina un modelo de bumerang de la forma que prefiera. Puede servir de orientación el que se representa en la figura de la izquierda y que corresponde a un bumerang con retorno diseñado por Herb Smith (si la persona que lo va a lanzar es zurda, deberá fabricarse un bumerang para zurdos, que es la imagen especular de uno para diestros). Coloque el modelo anterior sobre la plancha contrachapada y marque los bordes del mismo sobre la plancha con un lápiz.

Recórtese la forma del bumerang con una sierra de albardilla o con un serrucho. Dése luego, a los bordes y a la parte superior, la forma general mostrada en la ilustración. (Con la parte de abajo no se hace más que redondear lo que será el borde de ataque.) El borde de ataque debe ser romo y afilado el borde de salida, de suerte que la superficie superior posea una inclinación descendente para unirse con la superficie del plano inferior que no ha sido modificada.

Se sujeta la pieza recortada a un banco de carpintero con unas tuercas; con una escofina de superficie curva, hay que recortar y afilar los bordes y la parte superior. Se eliminan los estrías dejadas por la escofina y se termina de dar forma a la madera frotando sus superficies con papel de lija grueso colocado alrededor de un trozo de madera blanda. Antes de dar el acabado final a la superficie con otro papel de lija más fino, hay que probar el vuelo del bume-

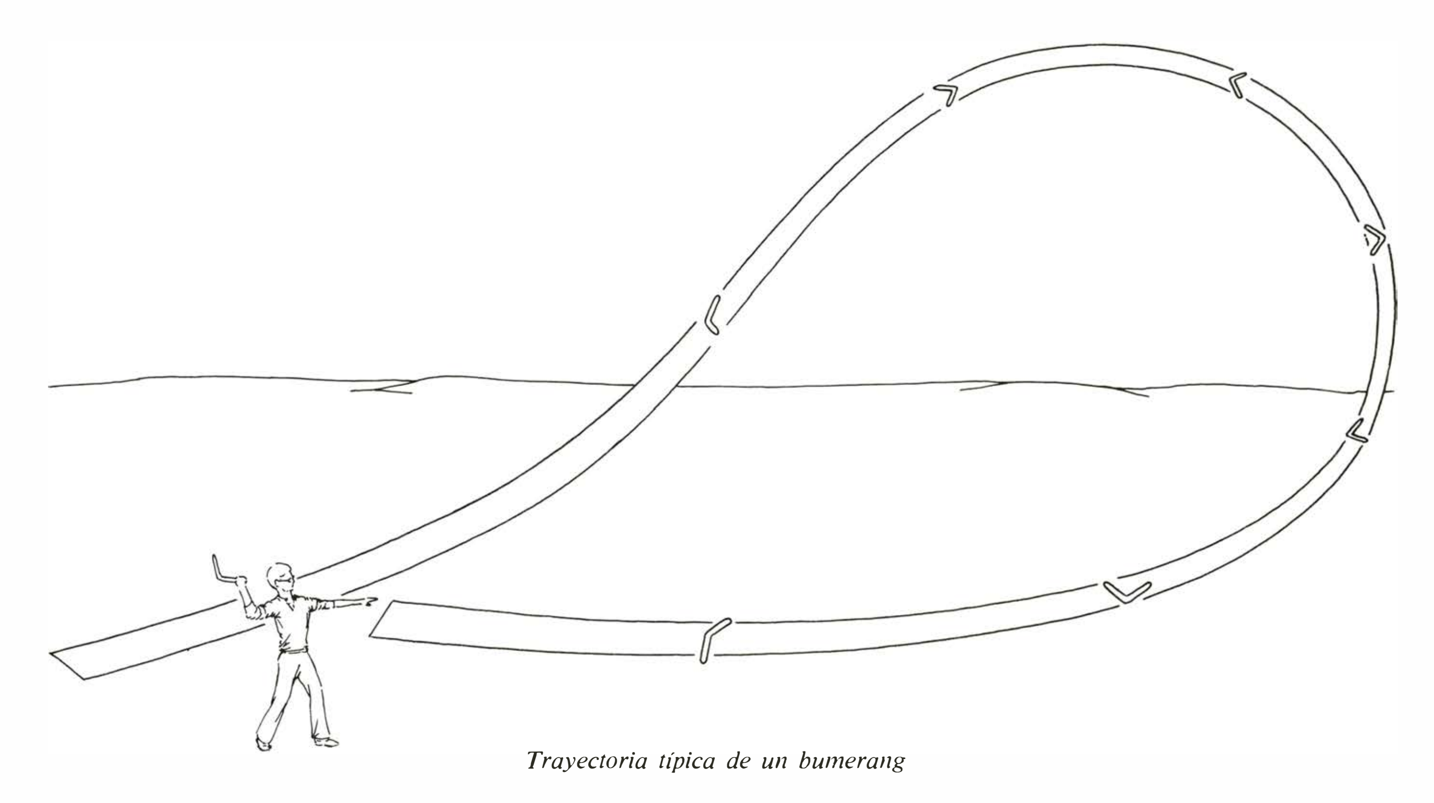


Disponiéndose para el lanzamiento de un bumerang

rang, a fin de que el ingenio pueda ajustarse bien con la escofina o bien con un papel de lija más grueso. Afinar un bumerang vale decir que primero se arroja, luego se recorta o se pule otro poco lanzándolo de nuevo y así hasta que el bumerang vuele en la forma pretendida.

Cuando hay mucho viento, el bumerang no puede volar bien. Si el viento es suave, debemos colocarnos frente a él, girar 45 grados a la derecha y lanzar el bumerang en esa dirección. Hay que mantener el aparato en posición vertical, sosteniéndolo por el extremo de uno de sus brazos (no tiene mucha importancia por cuál de ellos se sujete) con la cara plana hacia fuera. Lleve el bumerang por detrás de su cabeza, y láncelo entonces hacia el horizonte, procurando soltarlo cuando el brazo esté totalmente estirado hacia delante. No hay que pretender lanzarlo muy fuerte al principio; lo que cuenta es la manera de soltarlo, no la fuerza del lanzamiento. El bumerang permanece en el aire sin caerse debido al giro que se le ha proporcionado al impulsarlo.

La correcta orientación del juguete (es decir, el plano en que gira sobre sí

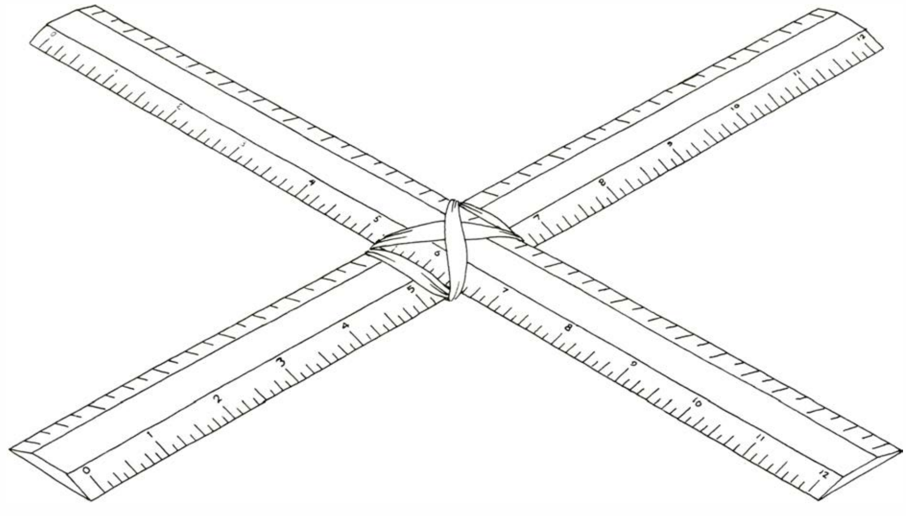


Trayectoria típica de un bumerang

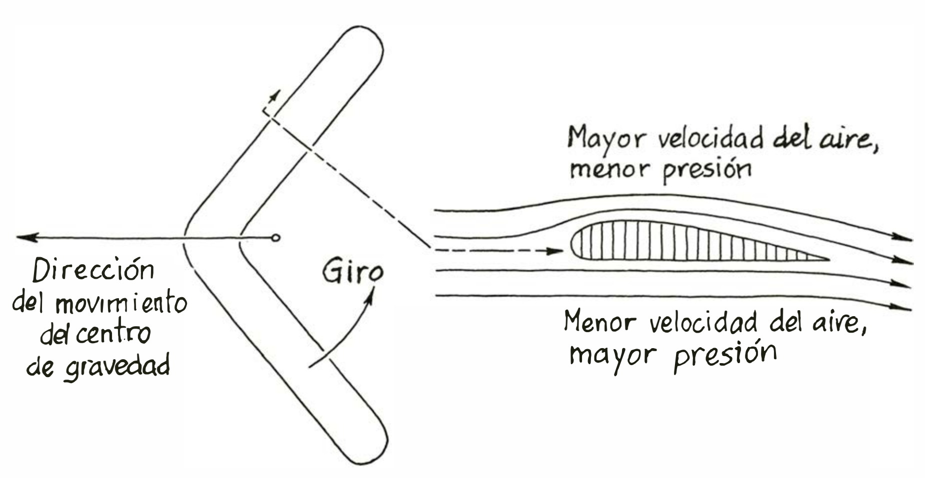
mismo) variará de acuerdo con las condiciones del viento y con el tipo de bumerang. Para conseguir un buen vuelo hay que lanzarlo con su plano de giro casi vertical. Bajo otras condiciones será necesario desviar el plano de giro (ladeando la parte superior del mismo hacia fuera) ángulos que pueden alcanzar los 45 grados. Mientras más desviemos de la vertical el plano en el que gira, mayor será la elevación inicial que realice el bumerang. Si el desvío inicial de lanzamiento es demasiado grande, puede elevarse demasiado rápido y desplomarse hasta el suelo con tal fuerza que puede incluso romperse.

En un vuelo correcto, el bumerang con retorno recorre en el plano horizontal una trayectoria circular alrededor de una esfera imaginaria. Cuando esté casi terminando el retorno, probablemente se remontará un poco y hasta llegará a trenzar un pequeño rizo antes de caer al suelo, a sus pies.

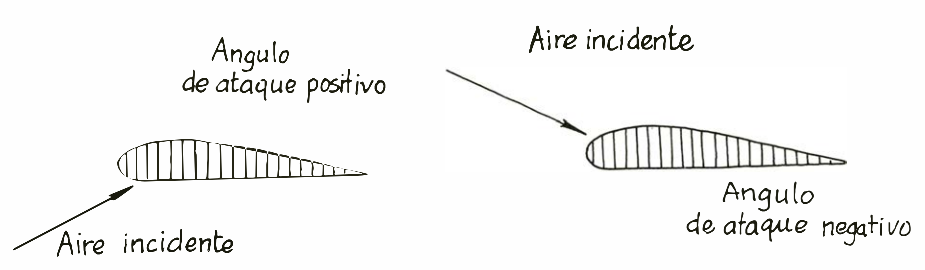
No obstante, si el lanzador tiene suerte, el bumerang puede hacer uno o dos círculos adicionales (menores que el primero) antes de caer. Aunque se lance el bumerang con su plano de giro casi vertical, probablemente retornará con el plano casi horizontal. Más adelante explicaremos por qué debe variar así el plano para que el bumerang complete su recorrido.



Un diseño para un bumerang de cuatro brazos hecho con dos reglas y una banda de goma



Sistema de corrientes de aire que rodean un brazo del bumerang en vuelo



Angulos de ataque positivos y negativos

Si a pesar de reinar un viento suave o ligero su bumerang cae sistemáticamente a su derecha, trate de lanzarlo un poco más a la izquierda de la dirección del viento. De igual manera, si el bumerang retorna demasiado a su izquierda, trate de lanzarlo más a la derecha de la dirección del viento. Si aterriza por detrás del lanzador, procurar impulsarlo con menos fuerza. Si esto no resulta, habrá que lanzarlo un poco por encima de la línea del horizonte con un plano de giro desviado un poco menos de la vertical. Si no hay viento y no consigue que el bumerang retorne totalmente, desvíe aún más el plano de giro de la vertical al objeto de ganar altura durante el vuelo.

Hay que procurar no dañar a nadie ni estropear nada con el bumerang. No debemos olvidar nunca su carácter peculiar de arma ofensiva. Se lanzará sólo cuando se esté en un gran espacio abierto. Si hay gente en los alrededores debemos asegurarnos de que saben que estamos ensayando un bumerang, con el fin de que estén atentas a esquivarlo en caso de que vaya en su dirección.

El buen ajuste de un bumerang depende, a la vez, de la experiencia y de la suerte. En general, si se le da una mayor curvatura a la superficie superior, volará más alto, pero el círculo que describirá a lo largo de su viaje será más cerrado. Aplanando más la superficie superior o curvando más la de abajo, conseguiremos que el bumerang vuele a menor altura, ya que la forma de la sección perpendicular de los brazos se asemejará menos a la clásica superficie de sustentación aerodinámica. Si la velocidad de rotación disminuye demasiado deprisa, de forma que el bumerang caiga al suelo a mitad de su vuelo, la explicación puede consistir en la excesiva fricción del aire sobre sus brazos, lo que le va restando velocidad de rotación. Aunque la superficie algo rugosa puede resultar beneficiosa para el vuelo, todas las rayaduras importantes originarán, casi con seguridad, corrientes adicionales de aire que acortarán el tiempo de vuelo.

No hay que empecinarse porque el bumerang tenga la forma típica ideal; quizá se prefiera doblarlo de manera que durante el vuelo el borde de ataque de cada brazo esté desviado a fin de deflectar el aire hacia la derecha, dando al bumerang una elevación hacia la izquierda. La existencia de este tipo de levantamiento es fácil de demostrar. Si se saca una mano por la ventanilla de un coche en marcha y se guía ésta de manera que tome diferentes ángulos con respecto a la corriente de aire, se notará

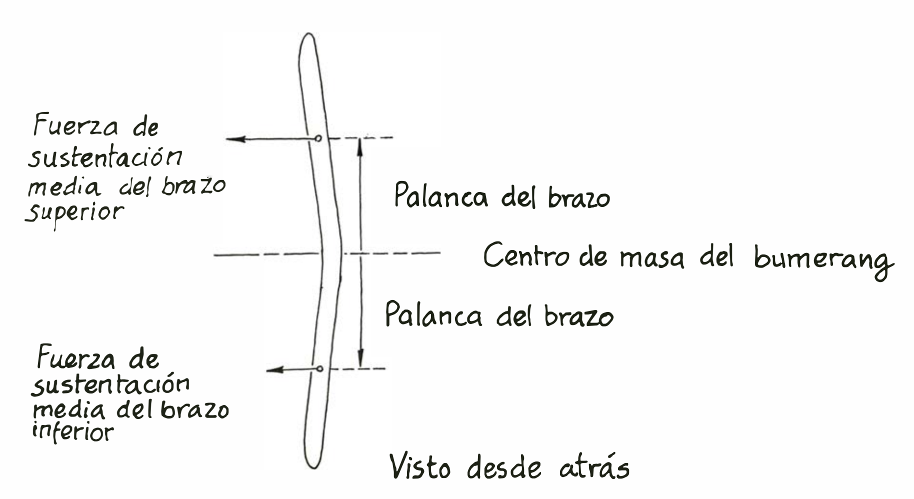
cómo la mano sufre un empuje ascendente. Para doblar un bumerang lo mejor es calentarlo de forma gradual por encima de los 200 grados centígrados y conferirle luego la forma deseada (por supuesto, con guantes) antes de que la madera se enfríe. Si lo hemos doblado demasiado, bastará calentarlo de nuevo para corregir el ángulo.

Una vez hemos construido un bumerang que vuele correctamente, podemos darle el acabado final forrándolo con papel convenientemente decorado. El excelente folleto de Smith que citamos en las referencias bibliográficas de la revista, explica cómo hacer este tipo de acabado y también da un gran número de diseños de bumerang.

Si se rompe el bumerang no tire los trozos, péguelos con epoxy, manteniéndolos juntos hasta que se sequen, puliendo a continuación la superficie hasta que vuelva a tener la forma deseada. Aunque el bumerang no será ya tan resistente como antes, posiblemente su nueva trayectoria de vuelo se habrá modificado, quizá de forma interesante, debido al cambio habido en la distribución de su masa como consecuencia de la ruptura y del arreglo.

El bumerang no tiene por qué limitarse sólo a dos brazos. Es más, uno de los bumerangs más sencillos de construir es el de cuatro hojas, ya que se puede hacer simplemente cruzando dos reglas iguales y uniéndolas por el centro. Este tipo de regla debe tener un poco curvada la superficie superior y bastante plana la superficie inferior. Podemos unirlos por el centro con un trozo de goma elástica resistente o con una tuerca y un tornillo aprovechando el agujero central que normalmente suelen traer. Este tipo de bumerang se lanza como los de dos brazos, pero hay que prestar atención a no cortarse con los bordes afilados y, sobre todo, no emplear reglas de bordes metálicos.

Se puede fabricar un bumerang sencillo en forma de cruz a partir de un cuadrado de cartón de unos 12-centímetros de lado. Para ello habrá que recortar un bumerang con tres o cuatro hojas, doblándolas un poco de manera que no todo el bumerang esté en el mismo plano. Se aumenta el alcance del bumerang añadiendo peso a sus brazos. A este modelo de cartón es muy sencillo aumentarle de peso colocándole clips en los extremos de sus brazos. Con él se pueden hacer demostraciones incluso en clase. Si su alcance es demasiado grande para las dimensiones del aula, se puede disminuir dicho alcance incrementando el doblez de los brazos de manera que el bumerang sea, en conjunto, menos plano, o bien doblando los brazos a lo largo de su línea central y en toda su longitud. Con esta última técnica, los brazos tienen una exagerada forma de superficie de sustentación: por una de las caras es tremendamente convexo, mientras que por la otra no. Como es normal, se debe lanzar el bumerang con la parte convexa hacia el que lo lanza. Modificando la forma de los brazos, desde una forma plana hasta lo más parecido a una superficie de sustentación, se puede elevar la altura de vuelo del bumerang, si bien a costa de que en su recorrido forme un círculo más cerrado.



Elevación media del brazo superior y del inferior

Cuando ya sepa lanzar bien el bumerang de madera, al lector se le puede ocurrir intentar atraparlo. El resultado puede ser un brusco golpe en los dedos. Pero si persiste en conseguirlo, conviene que mantenga las manos abiertas, enfrentadas entre sí y, cuando el bumerang esté revoloteando, girando sobre sí mismo en la última etapa de su vuelo, junte las manos rápidamente y procure agarrarlo en medio. Hay que tener cuidado en mantener los dedos fuera del giro de las aspas.

La explicación de por qué retorna un bumerang se basa principalmente en la sección perpendicular de sus brazos y en el giro que sobre sí mismo realiza el ingenio. Sin estas dos características, el bumerang se comportaría como cualquier otro palo. La forma de la sección perpendicular da al bumerang una ascensión aerodinámica del mismo tipo que la que producen determinadas formas de alas de avión. El giro sobre sí mismo le confiere estabilidad. Las leyes que rigen la mecánica de la rotación dan cuenta de cómo el movimiento giratorio hace que el eje sobre el cual se mueve el bumerang rote de una manera muy parecida a como el eje de giro de una peonza gira alrededor de la vertical. La fuerza ascensional y la estabilidad mantienen al bumerang arriba y la rotación del eje de giro le impulsa a retornar al punto de partida.

La fuerza ascensional aerodinámica puede explicarse con el sencillo ejemplo de la clásica ala de aeroplano, similar a la que describí en esta sección en abril de 1978 para explicar la ascensión de una cometa. La superficie de sustentación tiene el fondo plano, el frente romo, la parte de atrás afilada y la de arriba convexa. El aire pasa alrededor del ala más rápido por su parte superior que por su parte inferior. La razón se puede ver porque la corriente de aire tiene dos tipos de componentes. Un tipo de corriente fluye alrededor del ala sin rotaciones y con la misma velocidad tanto por arriba como por debajo del ala. La otra es una célula de circulación que fluye hacia atrás por la parte superior del ala y hacia delante por la parte inferior. Dicha circulación se forma en un ala real debido a la viscosidad del aire y su adhesión a la superficie del ala, que la obliga a seguir su forma incluso cuando ya se ha terminado por la parte de atrás la superficie curvada de la parte superior.

Cuando se superponen las dos corrientes ideales mencionadas, en la parte superior se suman y en la inferior se restan; resulta así que la velocidad del aire es mayor en la parte superior del ala que bajo ella. Esta diferencia es importante con respecto a la fuerza ascensional, ya que la presión del aire en una corriente es inversamente proporcional a su velocidad. Por tanto, la presión del aire es menor sobre el ala que bajo la misma: es decir, el ala sufre un empuje hacia arriba. (Un ala de un aeroplano real puede presentar una configuración de corrientes aerodinámicas más complicada que la de este sencillo modelo. Además, cuando un aeroplano viaja a gran velocidad, parte de la fuerza ascen-

sional puede venir del impacto del aire al pasar por debajo de un ala si ésta se encuentra ligeramente inclinada hacia arriba con el fin de deflectar el aire hacia abajo.)

Si una clásica superficie de sustentación está inclinada respecto a una corriente de aire, de manera que dicha corriente incida mucho sobre la superficie curvada de la parte superior, la fuerza ascensional es menor. Tal disposición se dice que tiene un ángulo de ataque negativo. En un modelo sencillo, la reducción de la fuerza ascensional obedece al empuje hacia abajo que produce la corriente al incidir sobre la parte superior. También se podría argüir que la ascensión se pierde parcialmente debido a que la tendencia del aire a rodear el ala es menor y, por tanto, la diferencia de velocidades del aire encima y debajo del ala también es menor.

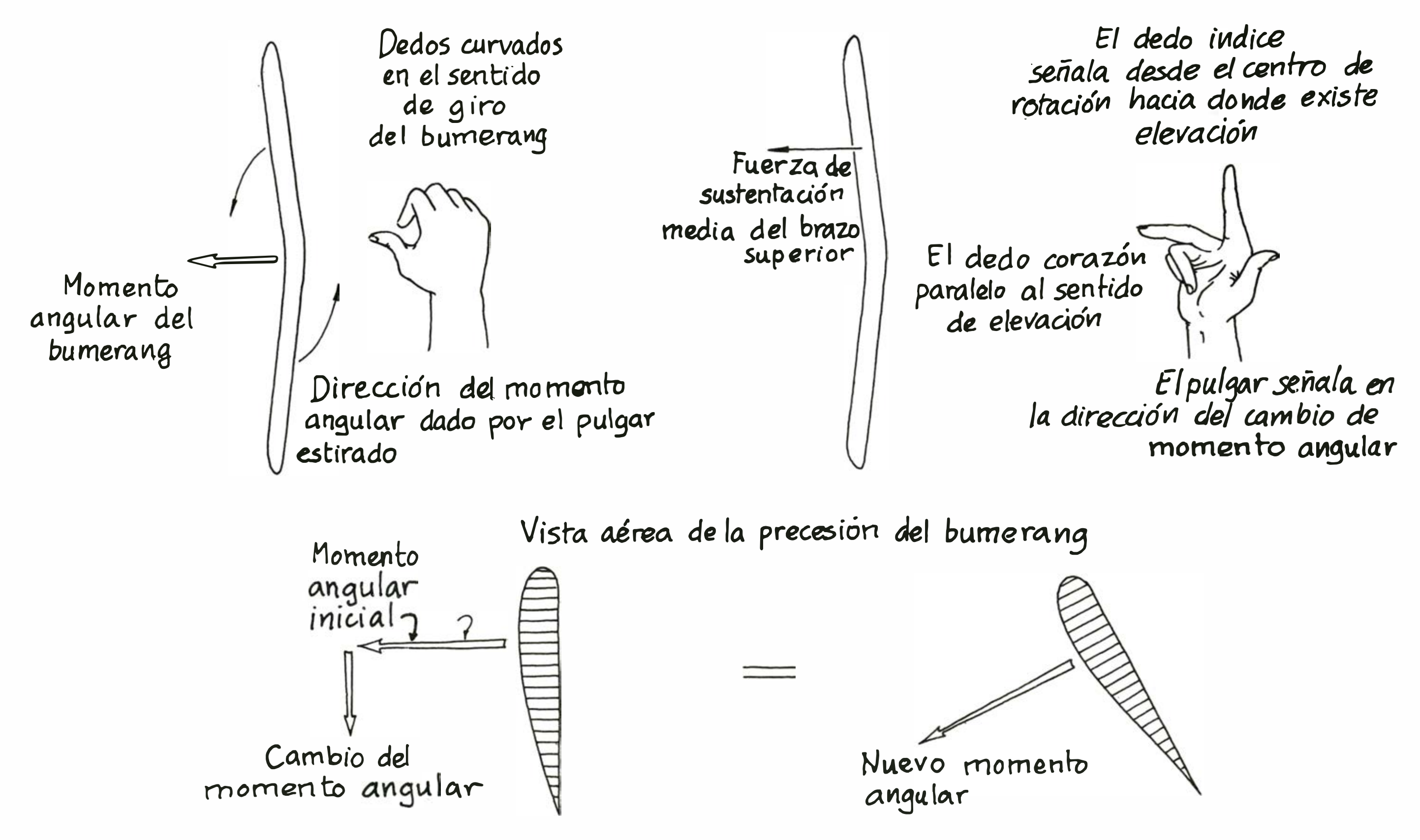
Por el contrario, si el ala se inclina de forma que la corriente de aire incida más en el lado plano bajo que en el alto, situación que sería llamada ángulo de ataque positivo, la elevación se incrementa en virtud del empuje ascensional procedente de la corriente de aire en la parte de abajo. También se incrementa la resistencia ofrecida por el aire. Si el ángulo es demasiado grande, la desventaja debida al incremento de la resistencia del aire es superior a la ventaja producida por la elevación. El ángulo de ataque de los brazos del bumerang cuando están girando en el aire es importante para su vuelo.

Los brazos de un bumerang pueden tener gran variedad de perfiles, pero la mayoría tienen una sección transversal similar a la de las alas clásicas. Normalmente, esta forma supone la existencia de un borde romo que gira en el aire a medida que el bumerang da vueltas y otro borde más afilado que es arrastrado durante el giro. Un lado suele ser plano y el otro convexo. Existen, sin embargo, numerosas variedades de esta estructura fundamental; parece que, hasta ahora, se han hecho pocos trabajos sistemáticos encaminados a determinar qué formas son las mejores desde el punto de vista aerodinámico. Algunos bumerangs son incluso planos por ambos lados, pero tienen los brazos torcidos al objeto de que el aire pueda deflectar a medida que éstos giran en el aire.

La fuerza de sustentación de un bumerang difiere bastante de la fuerza de sustentación de un ala clásica. Al comienzo de un vuelo la fuerza de sustentación del bumerang es, sobre todo, horizontal, con sólo la fuerza ascendente suficiente para mantener en equilibrio el peso del aparato. Puesto que el bumerang gira principalmente sobre un eje horizontal, los lados curvados de los brazos dan vueltas en un plano que es casi vertical, y la fuerza de sustentación casi horizontal. Por mor de simplicidad, en lo que sigue no tendremos en cuenta –por ahora– el peso del bumerang. Supondremos también que el bumerang es lanzado con la mano derecha, de manera que el plano de giro sea al principio exactamente vertical. La elevación se realizará hacia la izquierda del lanzador; así el bumerang empezará a moverse hacia la izquierda a medida que continúa girando en el plano vertical.

Si éste fuera todo el proceso, el bumerang no volvería nunca. Para ver por qué gira y regresa hay que entender primero qué otra influencia tiene la fuerza de sustentación sobre el bumerang. En particular importa saber cómo el par debido a la ascensión en el bumerang origina un movimiento de precesión del plano de giro.

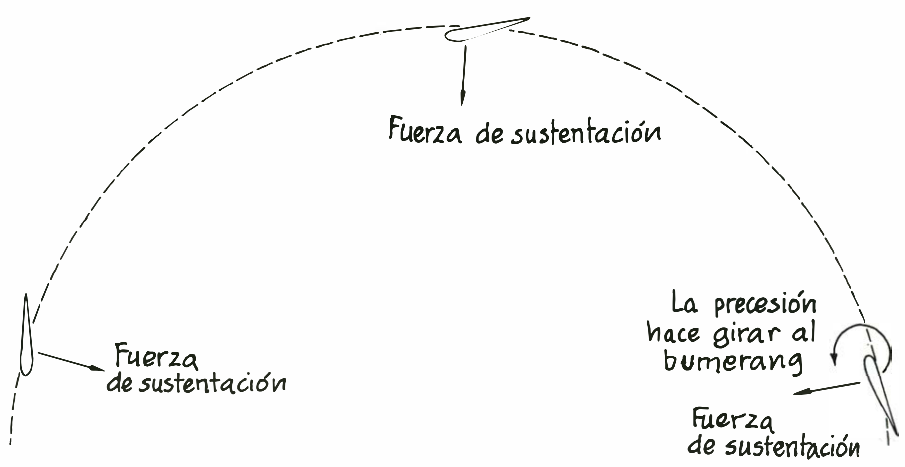
Imaginemos que uno de los brazos del bumerang ha ascendido en su giro hasta la posición más elevada posible y el otro brazo está casi en su posición más baja (estoy hablando del bumerang típico en forma de plátano). El brazo superior gira en la misma dirección que se mueve el centro del bumerang, mientras que el brazo inferior se mueve en dirección opuesta a la del centro. El aire próximo al brazo superior se mueve más de prisa (con relación al brazo) que el aire próximo al brazo de abajo. Por ello, se origina una fuerza de sustentación mayor en el de arriba que en el de abajo. La



Cómo determinar el momento angular y el cambio de éste

parte superior del bumerang experimentará siempre una mayor fuerza de sustentación y, por ello, un mayor impulso lateral que la parte inferior.

Mi primer pensamiento fue que la diferencia en la fuerza de sustentación (mayor fuerza en el brazo superior que en el inferior) harían que el plano de giro del bumerang se inclinara, dirigiendo así la fuerza resultante hacia abajo, lo cual tendría un efecto desastroso. Sin embargo, lo que realmente sucede es que la diferencia entre las fuerzas de sustentación produce una rotación del plano sobre un eje vertical. Esta rotación del plano de giro, conocida comúnmente como precesión, es la que hace que el bumerang retroceda.



Cómo mantiene la precesión el ángulo de ataque positivo

Para descubrir la causa de la rotación debemos examinar el par creado por la fuerza de sustentación. Tomemos el centro del bumerang como el eje sobre el cual éste gira (en realidad, el centro de masa alrededor del cual gira un bumerang con dos brazos es probable que esté bastante lejos de su punto medio, pero esto no invalida el argumento). Supongamos que el promedio de la fuerza de sustentación del brazo superior está dirigido horizontalmente hacia fuera desde el centro del brazo. De la misma forma, supongamos que el promedio de la fuerza de sustentación del brazo inferior está también dirigido horizontalmente hacia fuera desde el centro del brazo. El par creado por una de estas fuerzas, medido desde el centro del bumerang, es el producto de la fuerza de sustentación por la distancia hasta el punto en el que está aplicada, es decir, la mitad de la longitud del brazo. Dado que el brazo superior sufre un empuje ascendente mayor, también tiene un par mayor.

Si el bumerang no estuviera girando, la diferencia entre ambos pares provocaría sólo una inclinación de su plano. Por tener el brazo de arriba un par mayor, el plano se inclinaría en sentido contrario al que giran las agujas de un reloj visto por la persona que acabara de lanzar el bumerang. Sin embargo, el hecho de que éste gire supone una gran diferencia, porque el bumerang tiene entonces un momento angular y la tendencia a que el plano de rotación se incline se convierte en una rotación del mismo sobre el eje vertical.

El momento angular es el producto de la velocidad de rotación del bumerang por una función que relaciona la masa y la distribución de ésta en el sólido. Recurramos a un ejemplo de otro contexto: imaginemos que vamos a poner en marcha un tiovivo donde han subido varios niños. La fuerza tangente al borde que se aplica multiplicada por el radio del tiovivo es el par que le proporcionamos. Al empezar, el par origina una aceleración angular del tiovivo; la velocidad de rotación aumenta desde cero hasta algún valor final. ¿Cómo se deberían colocar los niños para conseguir una cierta aceleración angular con la menor fuerza posible? Instintivamente, los colocaríamos cerca del centro. Su masa es –por supuesto– la misma, pero su distribución con respecto al centro de rotación difiere. Cuando la masa se halla más cerca del centro, el tiovivo gira con más facilidad. La masa y su distribución son tenidas en cuenta por la función conocida como momento de inercia. Cuanto mayor sea la masa, o cuanto más lejos del centro se halle colocada, mayores serán el momento de inercia y la fuerza que habrá que aplicar para conseguir una aceleración angular dada.

Una vez que el tiovivo está girando y que ya no estamos empujando en el borde, el aparato tienen un cierto momento angular debido a su velocidad de rotación y a su momento de inercia. El momento angular se representa normalmente por un vector perpendicular al plano en el cual se mueve el objeto. Aquí el vector sería vertical. El sentido (hacia arriba o hacia abajo) se elige por convención como la dirección del pulgar de la mano derecha cuando la ponemos en pose de auto-stop con los dedos doblados en la dirección del giro del objeto.

El tamaño o el sentido de tal vector sólo se podrían modificar aplicando otro par al objeto. Con un tiovivo se podría volver a empujar en el borde otra vez. (A la hora de elegir cómo trazar un vector que represente el cambio del momento angular podría convenirse en dirigir el dedo índice de la mano derecha desde el centro de rotación hacia el lugar donde se aplica la fuerza y señalando con el dedo corazón en la dirección de la fuerza aplicada. Si se coloca el dedo pulgar de esa mano perpendicular a ambos dedos, éste señalará automáticamente en el sentido del cambio en el momento angular. El nuevo vector del momento angular es la combinación del anterior y del que representa el cambio.) Con un tiovivo al que se ha vuelto a empujar de forma tangente al borde, el nuevo vector seguiría siendo vertical, pero sería mayor o menor dependiendo de si la intención era hacer que éste girara más deprisa o más despacio.

Un bumerang que gire tiene dos pares actuando en los brazos: uno creado por el promedio de la fuerza de sustentación en el brazo superior y otro por el promedio de la fuerza en el brazo inferior. Dado que en el brazo superior es mayor, éste determina el momento angular y, por ello, ignoraré la fuerza sobre el brazo inferior (el razonamiento no cambiaría aunque incluyera la fuerza de sustentación del brazo inferior). Imaginemos al bumerang que se aleja del lanzador tras el impulso de éste con su brazo derecho. Gira en un plano vertical y tiene un vector momento angular dirigido hacia la izquierda del lanzador. La fuerza resultante en el brazo superior crea un par que cambiará la dirección del vector a medida que el bumerang continúa volando.

Para determinar cómo cambia el vector, podemos usar nuestra mano derecha, orientando el pulgar y el resto de los dedos adecuadamente. Con el dedo índice indicando desde el centro del bumerang al centro del brazo superior y el dedo corazón señalando hacia la izquierda, con el fin de estar en la dirección de la fuerza de sustentación en ese brazo, el pulgar estirado deberá apuntar hacia uno mismo. Así, el vector que representa el cambio del momento angular señalará en la misma dirección. La combinación mental del nuevo vector y del anterior se hace mejor mirándolo desde arriba. El vector que representa el

cambio es perpendicular al original y proporciona un nuevo vector desviado del anterior hacia el observador. El valor del momento angular no se altera porque el nuevo vector es perpendicular al anterior; sólo ha cambiado la dirección del momento angular, que ha girado sobre un eje vertical para apuntar más hacia el observador.

Este tipo de rotación de un vector de momento angular se denomina precesión; se puede observar cuando el eje de una peonza "precesa" en torno a la vertical. Otro ejemplo común de precesión se puede ver en el giro de una motocicleta. Las ruedas de una motocicleta giran lo suficientemente deprisa y tienen momentos de inercia lo bastante amplios como para hacer que su momento angular sea grande. Para que una motocicleta gire, no basta con torcer el manillar (como se haría con una bicicleta), sino que hay que inclinarla hacia el lado deseado. Los pares a que está sometida la motocicleta hacen "precesar" los vectores de momento angular de las ruedas y la motocicleta gira como un todo.

Durante la precesión del plano de rotación de un bumerang, éste continúa avanzando a una cierta velocidad, pero se deflecta continuamente debido a la fuerza horizontal que experimenta. El recorrido resultante es casi un gran círculo. En el vuelo con éxito de un bumerang el plano de giro "precesará" a la misma velocidad que lo hace el aparato en su recorrido. Su ángulo de ataque sigue siendo positivo. Esta proporción es necesaria para mantener los brazos con el ángulo de ataque adecuado.

Supongamos que el bumerang "precesa" demasiado despacio. En cuyo caso, a medida que éste se mueve a lo largo de su recorrido circular, su plano de rotación gira sobre un eje vertical a una velocidad más baja que la velocidad a la que se mueve el bumerang a lo largo de su recorrido. Cuando el plano de rotación se retrasa, el ángulo de ataque se va haciendo progresivamente negativo y el bumerang pierde altura.

Si el plano de giro "precesa" con demasiada velocidad, éste gira sobre un eje vertical más rápidamente que lo hace el bumerang como un todo viajando a lo largo del gran recorrido circular. En consecuencia, el ángulo de ataque se va haciendo progresivamente positivo hasta que el plano de rotación es perpendicular a la corriente de aire que recibe; mas, para entonces, la resistencia del aire ya habrá estropeado el vuelo.

El equilibrio entre la velocidad de precesión y la velocidad a que viaja el bumerang a lo largo de su gran recorrido circular no es crítico; en realidad es automático hasta cierto punto, ya que ambas velocidades dependen de la fuerza de sustentación. Es necesario lanzar repetidas veces el bumerang, cambiando cada vez la forma de las alas ligeramente hasta situarse cerca de este equilibrio y conseguir que regrese. Pero no conozco ninguna forma segura de corregir el vuelo de un bumerang que ya ha fracasado persistentemente.

El recorrido circular del bumerang es independiente de la velocidad con que se lanza. El momento de inercia y la forma de sección transversal del mismo determinan el radio del recorrido. Así pues, con un bumerang dado se podrá conseguir un mismo círculo grande (con el mismo lanzamiento de éste en el plano vertical que hemos supuesto) con independencia de la fuerza con que se lance el aparato (siempre que se lance lo suficientemente fuerte como para que pueda completar su recorrido). Si se desea cambiar la amplitud del círculo, hay que elegir un bumerang distinto con un momento de inercia diferente o de otra sección transversal. El mes próximo explicaré cómo se puede añadir lastre a los brazos con el fin de incrementar su momento de inercia. Esta técnica se usa por los lanzadores de bumerangs en su intento de batir records de distancia.

El "frisbee" vuela de forma bastante parecida. Tiene una superficie superior curvada y se lanza con una sacudida de la muñeca para hacerle girar. El "frisbee" gana altura por medio del impacto del aire o por la diferencia en la velocidad del aire entre su parte alta y la baja. Un "frisbee" debidamente lanzado en un plano casi vertical volverá al lanzador lo mismo que lo hace un bumerang. Sin embargo, lo normal es que un "frisbee" se lance de modo que gire ligeramente hacia otra persona, y así el lanzador orienta el plano de giro para proporcionarle justo el impulso horizontal suficiente para que pueda describir la curva.

Tanto el bumerang como el "frisbee" pueden ser deslizados a ras del suelo sin destruir el vuelo. Imaginemos un "frisbee" que pase rozando el suelo con su borde guía inclinado ligeramente hacia abajo. Este borde golpeará el suelo; la fuerza del suelo en el punto de contacto origina un par en el aparato y cambia el momento angular, pero debido a que el vector cambio es casi perpendicular al vector momento angular original, el nuevo vector momento angular es sólo una rotación del original. El momento angular no cambia significativamente de tamaño, sólo modifica el sentido. El giro del "frisbee" no se reduce mucho, el aparato sólo es reorientado y sigue deslizándose en otra dirección.

Libros

La naturaleza humana, sus posibilidades fisiológicas, sociología de la medicina y quimioterapia del cáncer

R. Margalef, J. Tamarit, L. Vila y A. González

On Human Nature, por Edward O. Wilson. Harvard University Press; Cambridge, Mass. & London, 260 págs; 1978. El autor de este libro, Prof. E. Wilson, es un biólogo, ilustre principalmente por sus investigaciones sobre los insectos sociales, que en 1975 publicó un volumen titulado *Sociobiology, the new synthesis,* que, para mi gusto, fue el mejor libro de biología aparecido dicho año. En él Wilson intentaba un estudio comparado de las formas de comportamiento, heredadas o aprendidas, que constituyen el fundamento de los llamados fenómenos sociales, en los animales y en el hombre. El punto de vista adoptado era muy razonable y estaba en la tradición darwinista de suponer que el hombre no estaba separado del resto de la naturaleza por una diferencia cualitativa. Aceptando que una gran parte del comportamiento humano se basa en el aprendizaje y en la tradición cultural, consideraba que su infraestructura era, por supuesto, hereditaria y, como tal, sometida a la selección natural. Esta interpretación corresponde a la ortodoxia biológica del momento. Pero, para ciertos círculos intelectuales modernos, es anatema, como lo era hace medio siglo para ciertos círculos españoles la teoría de la evolución, pura y simple. Se podría anticipar que el autor sería maltratado de materialista y ateo por unos y de racista y aun de fascista por otros, y estos últimos han sido los más vociferantes. La sociobiología ha levantado polémicas que han contribuido a vitalizar y dar nombre a la universidad de Harvard, sede de los más notorios contendientes.

En este nuevo libro, Wilson trata de resumir sus puntos de vista en lo que se refiere al hombre y aplacar a los que le han puesto el sambenito de reaccionario. No creo que lo consiga y esta obra, al tratar del hombre en un contexto mucho menos amplio, pierde la emoción de los grandes paisajes que tenía el libro anterior. Aquí la evolución se usa como el marxismo, o la psicología, o cierta ecología, para explicarlo todo *a posteriori* y lo que se trata de comprender y perdonar es un catálogo de problemas humanos que encabezan los capítulos de la parte más sustanciosa del libro. Agresión, sexo (con una buena parte dedicada a la homosexualidad, cómo no), altruismo, religión y esperanza, son los epígrafes bajo los cuales se ofrecen y comentan muchos datos interesantes, a la luz de una filosofía que no me parece que supere mucho a la de los "Welträtsel" de Haeckel, a pesar de los años que han pasado. El libro se lee con gusto, porque refleja con vivacidad polémicas apasionantes y actuales y está bien escrito y documentado. Más, para muchas sensibilidades, entre otras la mía, la proyección humanística de la ciencia ha de ser menos "explicar" que sentir simpatía por la aventura humana. Al fin y al cabo, lo que se reprocha a la ciencia es la incongruencia del observador "desapasionado" embarcado en la misma nave y partícipe del destino común de la humanidad. Wilson se da cuenta de ello, pero su afán de presentarse como progresista le deja un poco envarado en un claroscuro donde la cordialidad, la autocrítica y el buen humor serían los mejores antídotos contra los dogmatismos. (R. M.)

Fisiologia Ambiental y Espacial. Dirigida por Augusto Corominas.– EUNIBAR–Editora Universitaria de Barcelona. Segunda edición 1978. Se trata de un libro de 244 páginas, escrito por un grupo de especialistas de la fisiología humana en ambientes especiales, de reputación merecida en sus respectivos campos y coordinados por el profesor Augusto Corominas Vilardell. Está dirigida a los estudiantes de medicina y su objeto es el estudio del efecto de las variaciones ambientales sobre las funciones fisiológicas, especialmente del hombre. Este objetivo se toma como definición de fisiología ambiental en los dos primeros capítulos. En el primero se destaca la necesidad de considerar al hombre como un sistema integrado, como una unidad anatómica, funcional y psíquica. La unidad es una característica del ser en general, destacada especialmente por Ludwig Bertalanffy para los seres vivos, que deben ser considerados como organismos, es decir, como "todos" cuyas propiedades no pueden deducirse por simple suma de las que tienen sus partes. En el capítulo I, por medio de esquemas adecuados se resume la relación de la fisiología con otras ciencias del medio ambiente y los distintos factores del nicho ecológico que actúan sobre el hombre. Cabría considerar una superunidad del hombre con su medio, pero los autores se limitan voluntariamente a estudiar las acciones del medio sobre el hombre y excluyen el estudio de las modificaciones del medio por efecto de la vida.

Reconociendo la importancia del mantenimiento de la constancia del medio interno, señalada por Claude Bernard y denominada homeostasis por Cannon, el hombre se encuentra en su vida, en un punto concreto del planeta, con variaciones en el medio ambiente que se suceden según distintos ritmos, bien definidos y enumerados en el capítulo II, dedicado a la cronobiología; este capítulo es de mucha importancia en relación con la existencia de un tiempo biológico, propiamente dicho. Las figuras y esquemas de esta parte son muy valiosas y sería deseable una explicación de los diagramas polares y una mejor definición de conceptos, como el de acrofase. Es ingenua la afirmación de que la cronobiología es una ciencia que se basa en la estadística y que obliga al uso de ordenadores, pues estas características son aplicables a todas las ciencias de la naturaleza; conviene concretar especialmente las circunstancias en que es casi obligatorio el empleo de computadores.

En el capítulo III se estudia los mecanismos fisiológicos de adaptación del organismo humano a situaciones ambientales extremas. Se dividen en tres apartados: en el primero se estudia el agua, en el segundo el calor y, en el tercero, la altura. El apartado sobre el agua es excelente desde el punto de vista hidrográfico y se echa de menos un resumen de la tectónica de placas, expuesta con la misma lucidez. Sin embargo este apartado es poco biológico, puesto que se omite la influencia del agua de bebida y su contenido en algunos elementos (calcio, fluor, iodo, etcétera) para la fisiología humana. En el capítulo IV, dedicado a la fisiología del deporte, se expone brevemente la metodología para obtener los perfiles biológicos y psicológicos del deportista, lamentando acertadamente la falta de coordinación entre medicina del deporte y fisiología pura en nuestro país.

En los capítulos siguientes, se estudia

el efecto de ambientes extremos: bajas temperaturas (capítulo V), explicando las precauciones y tratamiento de la congelación; la influencia del ruido (capítulo VI), con una introducción sobre la audición, concepto y clases de ruidos, sus efectos sobre el organismo humano y su profilaxis. La vida en ambientes profundos subterráneos se trata en el capítulo VII, y en profundidades submarinas en el capítulo VIII, considerándose los efectos del O_2 comprimido, los efectos de la anoxia y la enfermedad descompresiva. En un grupo de capítulos relacionados se trata en forma excelente la fisiología aeronáutica (capítulo IX), la fisiología del cosmonauta (capítulo XI) y del paracaidista (capítulo X), y la vida en el espacio (capítulo XII).

En el capítulo XIII se hace una especie de recapitulación de la base científica de la relación del hombre con el medio, que es de gran utilidad. Se hace una consideración general de los flujos de materia y energía, con los grandes ciclos (C y N); se detallan los parámetros físicos esenciales del nicho ecológico (presión, temperatura, O_2, agua, alimentos, aceleración gravitatoria, etcétera), las radiaciones cósmicas y la contaminación atmosférica. Se trata globalmente el fenómeno de la adaptación, sus limitaciones, sus mecanismos y su evolución filogenética.

En el capítulo XIV, final del libro, se hace una descripción de la influencia ambiental sobre la respuesta a los fármacos. En resumen, la obra comentada tiene objetivos ambiciosos que cumple satisfactoriamente y puede recomendarse no sólo a estudiantes de medicina, sino también a personas de cultura media adecuada que se interesen por los efectos del medio ambiente sobre la vida humana. (J. T.)

SOCIOLOGIA DE LA MEDICINA. UNA INTRODUCCION CRITICA, por Jesús M. de Miguel; Editorial Vicens Vives; Barcelona, 1978, 294 págs. Por diversas circunstancias, la sociología de la medicina está poniéndose de moda entre nosotros. Jesús de Miguel añade, con la presente obra, un título más a esta bibliografía sanitaria española, cuya necesidad era evidente. El aspecto sanitario, los servicios públicos de salud, las demandas percibidas en este sector; todo ello ha entrado en el horizonte de las preocupaciones nuestras de cada día, que han asumido los partidos políticos a la hora de configurar sus proyectos de política social. Interesante momento, pues, para quien ofrezca información, metodología y diseños en torno a la salud, en el contexto global de una sociedad de cambio.

Pero dado que no existe un acuerdo acerca de cómo lograr la unión de la investigación sociológica con la práctica médica y sanitaria, no hay que perder de vista los enfoques peculiares de la sociología y la medicina. El sociólogo puede considerar a la medicina y la salud como "ámbitos" en los cuales hacer la aplicación de su teoría sociológica general. Pero también puede aplicar sus investigaciones sociológica a la solución de problemas médicos. La medicina invita a la sociología a adentrarse en zonas de especial interés médico, entendiendo la aportación sociológica bien al nivel de experiencia metodológica trasladable al campo sanitario o requiriendo teorías sociológicas como explicativas de problemas médicos.

Si existen diferencias entre la cultura médica y la sociológica, una obra que intenta ser puente entre ambas puede no satisfacer ni a médicos ni a sociólogos.

Suponemos que no será el caso de esta obra, que ha superado la prueba de servir como texto de sociología de la medicina para alumnos de medicina de la Universidad Autónoma de Barcelona. Esto, que es una garantía, constituye también una limitación: tenemos en las manos un "libro de texto", una "introducción crítica", que "no cubre todos los temas posibles de la sociología de la medicina". Con este criterio, la obra aparece dividida en cinco partes. "En la primera se analizan los *paradigmas teóricos* de la disciplina, desde el obligado ¿qué es la sociología de la medicina? hasta los diversos modelos que se utilizan para el análisis de lo que sea salud. En la segunda parte se profundiza en algunos aspectos de la *micro-sociología de la medicina,* es decir, aquellas relaciones más individuales ya sean de la persona humana consigo misma, o con otros seres humanos. La tercera parte incide en la *macro-sociología de la medicina* analizando grupos secundarios, organizaciones complejas, clases sociales e incluso estudiando el sistema sanitario global. La cuarta parte recoge algunos de los *aspectos metodológicos* más importantes de esta disciplina para entender su desarrollo. En la quinta parte se estudian con mayor detalle las relaciones entre *cambio social y sanidad*" (p. XV).

Quiero ahora comentar algunos aspectos que trata De Miguel y que se sitúan en el centro de las discusiones actuales sobre estos temas.

La naturaleza de las relaciones médico-enfermo, las características del "acto médico", no solamente oponen las posturas de Parsons y Goffman, demostrando o refutando el enfoque estructural funcional. Para cualquiera que esté al tanto de la prensa profesional, las discusiones sobre el acto médico y la relación médico-enfermo están enfrentando los dos enfoques de la organización sanitaria (socialización-liberalización) y los partidos políticos, o mejor, las ideologías que los inspiran. La lectura de las páginas 71-81 ayudan a situar este problema. Pero faltaría una ampliación que con sentido crítico se ocupase de los argumentos en torno a la humanización del acto médico que no tiene por qué ser privativa del enfoque liberal-privado de la medicina. La confianza, como elemento destacado de la relación médico-enfermo, es aplicable igualmente a la medicina en equipo, por alto que sea el nivel tecnológico en que se desarrolle. Es un problema de educación sanitaria que se vincula al rol que el paciente —y en general, la cultura— asigna al médico. Y lo mismo dígase del tema de la libre elección de médico que, a determinados niveles, nadie niega. En la controversia actual sobre los enfoques de la medicina, de cara a una reforma sanitaria, mucho se esclarecería el camino.

Otro ejemplo: en una reciente encuesta a la clase médica, ésta se pronunciaba mayoritariamente por una defensa profesional —y no sindical— de sus intereses. Aparte de la extendida alergia al asociacionismo sindical, deberíamos preguntarnos cuál es el esquema de relaciones de poder e intereses entre la sociedad española y las asociaciones profesionales médicas y sanitarias. Según De Miguel, poco hay escrito (pág. 85) pero, añado, mucho se va a poder escribir partiendo de las investigaciones llevadas a cabo por el Ministerio de Sanidad, los Colegios Médicos y las secciones de sanidad de los partidos y centrales sindicales. La figura del sanitario en un proceso de cambio social es, sin lugar a dudas, la del profesional más afectado por las tendencias socializadoras en la prestación de los servicios de salud y, consiguientemente, indica el campo en el que los grupos de presión son más activos y poderosos (pp. 85-113).

Fijando la atención en los modelos sanitarios de los partidos políticos encontramos que el público no los conoce y, paradójicamente, los profesionales sanitarios no siempre están al tanto. La mitad de los médicos —según datos de la aludida encuesta— confiesan su ignorancia de esos programas. El autor se esfuerza en el cap. 6 (pp. 115-138) en presentar las posibles respuestas de cada partido a los problemas fundamentales del sistema sanitario español en la actualidad. Son previsiones de futuro, protagonizadas por cuatro bloques políticos: (comunista, socialista, democracia cristiana y frente nacional) y que reflejan las posturas de éstos ante cuestiones como

la cobertura y regionalización de los servicios sanitarios, la coordinación de la asistencia sanitaria, la articulación de los sectores público y privado, el control de la industria farmacéutica, la financiación de la sanidad, el control de la organización sanitaria y la realización de la reforma sanitaria.

Con el enfoque de la teoría de sistemas, el autor aborda en el cap. 7 la construcción de un modelo explicativo del sistema sanitario español. Poniendo de relieve las relaciones causales básicas entre los diferentes subsistemas explica la evolución del ejemplo español con conclusiones indudablemente válidas. El tono determinista y automático que tanta brillantez da a los modelos, debiera estar contrapesado por otras aportaciones teóricas del pensamiento sociológico.

Por ejemplo, cuando el autor mantiene la tesis de que en una primera etapa de desarrollo las diferencias en los niveles de salud de la población dependen más del propio desarrollo económico que de la organización sanitaria, siendo la programación y planificación, conforme a un modelo coherente, la exigencia que en una etapa posterior se requiere para seguir progresando. En este momento cuando la decisión política tiene lugar, es necesario deslindar lo técnico y lo ideológico en los motivos para la adopción de decisiones. La posibilidad de camuflar con seudo razonamientos lógicos (utilizando especialmente la lógica de los números y los porcentajes) sentimientos profundos, ideologías cargadas de valor y adhesiones primarias no racionales, es un tipo de acción social muy extendida. Pienso que una vuelta a la sociología de Pareto sería ventajosa. Sus conceptos de "residuos" y "derivaciones" constituyen un instrumental válido para la comprensión de la acción social y los razonamientos que la justifican. La capacidad crítica ante la propaganda política, más o menos disfrazada, se robustecería. Otro problema es si de aquí nace el escepticismo desilusionado.

Las decisiones políticas son inevitables ante la reforma sanitaria, que mayoritariamente se extiende como una necesidad inaplazable. Los intereses que existen antes, en y después de la misma, sólo se pueden conocer accediendo a la historia de las "reformas que fueron", empezando por los Planes de Desarrollo. Frente a la tentación de "reorganizar para no cambiar nada", las expectativas ante el futuro presentan hoy —es mi opinión— un importante acopio de información estadística sobre las que incidirán los proyectos políticos hoy existentes. Proyectos que, probablemente, no todos serán del gusto del autor, pero que en cierto sentido son coherentes con las filosofías políticas de los partidos que se disputan el favor del electorado. Y éste somos todos.

Quiero terminar citando las cuestiones que para De Miguel resumen los objetivos de una reforma sanitaria para España. Estos objetivos se contemplan en los proyectos de reforma sanitaria. Pero su grado de consecuencia, lo mismo que la relación entre los mismos, es aún incógnita a despejar. Ya es importante la coincidencia en los siguientes puntos: 1) un servicio cuyo nivel de cobertura se corresponda con toda la población; 2) formas de unificación de la medicina preventiva, curativa, rehabilitativa y de servicios; 3) articulación de los restos de la medicina privada en el conjunto del sistema sanitario; 4) control de los medicamentos y papel de la industria farmacéutica; 5) fiscalización del coste de la sanidad; 6) control de los diversos niveles sanitarios y democratización de la organización sanitaria; 7) por último, el desafío que supone encontrar una política eficaz para contrarrestar el poder de los grupos de presión implicados (pp. 285-287).

A lo largo de este comentario acerca del contenido de la "sociología de la medicina" he querido destacar aquellas cuestiones que pueden resultar más interesantes para el público más amplio que lea provechosamente esta obra, dejando de lado capítulos más complejos para los que se requieren especial preparación médica, estadística o sociológica. En un país como el nuestro hay dos cuestiones —salud y enseñanza— cuyo tratamiento exige imaginación más que simple traslado de recetas foráneas. La obra de Jesús M. de Miguel es una aportación que invita a pensar. (L. V. L.)

QUIMIOTERAPIA DEL CANCER, por Martín J. Cline y Charles M. Haskell. Editorial AC. 363 páginas; Madrid, 1977. Este pequeño libro escrito con las clásicas virtudes de los libros norteamericanos, bien sistematizados, sencillos y sin grandes pretensiones, pero que llevan dentro de sí un trabajo excelentemente realizado, sintético y fundamental, está concebido en diecisiete capítulos, que a su vez pueden agruparse en dos partes. Una primera general, dedicada a una introducción a la quimioterapia con una breve revisión histórica, una revisión de los factores que modifican los efectos y una pequeña referencia a los diferentes agentes y una sucinta relación farmacocinética. La segunda parte está formada por los principales grupos nosológicos. Por último, una sección final dedicada a las complicaciones y al futuro de la quimioterapia. La estructura de cada grupo nosológico consta de una pequeña introducción a la patología y clasificación tumoral, con una amplia valoración de las terapéuticas actuales y una extensa bibliografía, que considero de un valor capital para el desarrollo del libro y de lo que él puede poner en marcha.

Dentro de los grupos nosológicos destacan por su importancia los del carcinoma de mama, leucemias, linfomas y mieloma, por ser quizá donde la quimioterapia obtiene unos resultados más esperanzadores. Quizás en este aspecto y recordando que la concepción del libro y de la colección donde va inserto es un planteamiento básico, encontramos que la filosofía que informa las pautas de tratamiento tiene ciertos tintes conservadores, aunque esta observación no debe entenderse como juicio adverso sino como expresión de la identidad de un libro y de la intencionalidad de los autores. Estos conceptos vienen muchas veces explicados sobre un caso ejemplo, a la vez que por unos esquemas muy simples que pueden considerarse esbozos de protocolos de tratamiento. En aspectos considerados como conflictivos, el libro adopta una inteligente posición ecléctica: las verdades absolutas y los dogmatismos en medicina cada vez son más escasos.

En el desarrollo de los temas existe un concepto repetido en muchas ocasiones y que tiene un gran valor didáctico. Me refiero al tratamiento multimodal de los enfermos cancerosos que se vislumbran en los sucintos esquemas que, a modo de guía del tratamiento, aparecen en casi todos los capítulos.

El excelente fondo bibliográfico que aporta, la simplicidad y claridad de conceptos dan a este libro un valor singular para todos los estudiantes de los últimos cursos de medicina y en especial para los postgraduados en formación oncológica que serán los que más partido podrán obtener de sus páginas.

Se trata de un libro editado en 1977. Aunque el tiempo ha podido variar algunos conceptos si bien los fundamentales continúan vigentes, nos parece adecuado que siente sus bases sobre hechos e ideas ampliamente aceptadas, dejando las novedades experimentales para las publicaciones periódicas. En este sentido, el libro, muy ameno por otra parte, deja abierto el camino; consciente de su temporalidad, sólo intenta resumir y concentrar el estado de la quimioterapia en el momento en que se escribió. (A. G. N.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

OBJETOS APOLO

METEORITES AND THE ORIGIN OF PLANETS. J. A. Wood. McGraw-Hill Book Company, 1968.

METEORITES: CLASSIFICATION AND PROPERTIES. John T. Watson. Springer-Verlag, 1974.

SPACE GEOLOGY: AN INTRODUCTION. Elbert A. King. John Wiley & Sons, Inc., 1976.

THE ASTEROIDS. Clark R. Chapman, James G. Williams y William K. Hartmann en *Annual Review of Astronomy and Astrophysics*, vol. 16, págs. 33-75; 1978.

METASTASIS CANCEROSAS

A PATHOBIOLOGIC OVERVIEW OF METASTASIS. L. Weiss en *Seminars in Oncology*, vol. 4, n.º 1. págs. 5-19; marzo, 1977.

EXPERIMENTAL TUMOR METASTASIS: CHARACTERISTICS AND ORGAN SPECIFICITY. Garth L. Nicolson en *Bioscience*, vol. 28, n.º 7, págs. 441-447; julio, 1978.

SPECIFICITY OF ARREST, SURVIVAL, AND GROWTH OF SELECTED METASTATIC VARIANT CELL LINES. Garth L. Nicolson y Kennteh W Brunson en *Cancer Research*, vol. 38, n.º 11, págs. 4105-4111; noviembre, 1978.

EL ESPECTRO DEL HIDROGENO ATOMICO

OPTICAL RESOLUTION OF THE LAMB SHIFT IN ATOMIC HYDROGEN BY LASER SATURATION SPECTROSCOPY. T. W. Hänsch, I. S. Shahin y A. L. Schawlow en *Nature Physical Science*, vol. 235, n.º 56, págs. 63-65; 24 de enero de 1972.

AN EXPERIMENTAL DETERMINATION OF THE RYDBERG CONSTANT. B. P. Kibble, W. R. C. Rowley, R. E. Shawyer y G. W. Series en *Journal of Physics B*, vol. 6, págs. 1079-1089; junio, 1973.

HYDROGEN 1*S*- 2*S* ISOTOPE SHIFT AND 1*S* LAMB SHIFT MEASURED BY LASER SPECTROSCOPY. S. A. Lee, R. Wallenstein y T. W. Hänsch en *Physical Review Letters*, vol. 35, n.º 19, págs. 1262-1266; 10 de noviembre de 1975.

LA PERCEPCION DE LOS BLANCOS Y NEGROS DE LAS SUPERFICIES

DISCOUNTING THE BACKGROUND–THE MISSING LINK IN THE EXPLANATION OF CHROMATIC INDUCTION. J. Walraven en *Vision Research*, vol. 16, págs. 289-295; 1976.

PERCEIVED LIGHTNESS DEPENDS ON PERCEIVED SPATIAL ARRANGEMENT. Alan L. Gilchrist en *Science*, vol. 195, n.º 4274, págs. 185-187; 14 de enero de 1977.

LA TEORIA RETINEX DE LA VISION DEL COLOR. Edwin H. Land en *Investigación y Ciencia*, n.º 17, págs. 64-81; febrero, 1978.

CANALES IONICOS EN LA MEMBRANA DE LA CELULA NERVIOSA

ORGANIZATION OF THE IONIC CHANNELS IN NERVE MEMBRANES. Richard D. Keynes en *The Nervous System, vol. 1: The Basic Neurosciences*, dirigido por Donald B. Tower. Raven Press, Publishers, 1975.

IONIC CHANNELS IN EXCITABLE MEMBRANES: CURRENT PROBLEMS AND BIOPHYSICAL APPROACHES. Bertil Hille en *Biophysical Journal*, vol. 22, págs. 283-294; 1978.

DELFINES

MAMMALS OF THE SEA: BIOLOGY AND MEDICINE. Dirigido por Sam H. Ridgway. Charles C. Thomas, Publisher, 1972.

OCURRENCE AND GROUP ORGANIZATION OF ATLANTIC BOTTLENOSE PORPOISES *(TURSIOPS TRUNCATUS)* IN AN ARGENTINE BAY. Bernd Würsig en *The Biological Bulletin*, vol. 154, n.º 2, págs. 348-359; 1978.

CATAPULTAS ANTIGUAS

THE MECHANICAL TECHNOLOGY OF GREEK AND ROMAN ANTIQUITY. A. G. Drachmann. E. Munksgaard, 1963.

GREEK CATAPULTS AND CATAPULT TECHNOLOGY: SCIENCE TECHNOLOGY AND WAR IN THE ANCIENT WORLD. Barton C. Hacker en *Technology and Culture*, vol. 9, n.º 1. págs. 34-50; enero, 1968.

GREEK AND ROMAN ARTILLERY. E. W. Marsden. Oxford at the Clarendon Press. 1969-71.

GEARS FROM THE GREEKS: THE ANTIKYTHERA MECHANISM– A CALENDAR COMPUTER. Derek de Solla Price en *Transactions of the American Philosophical Society, New Series*, vol. 64, parte 7; 1974.

INMUNOHISTOQUIMICA EN LA MICROSCOPIA ELECTRONICA

ENZYME-LABELED ANTIBODIES: PREPARATION AND APPLICATION FOR THE LOCALIZATION OF ANTIGENS. P. K. Nakane y S. B. Pierce, en *J. Histochem. Cytochem.* 14: 929; 1966.

INTRACELLULAR LOCALIZATION OF BASEMENT MEMBRANE ANTIGEN IN PARIETAL YOLK SAC CELLS. A. Martínez-Hernández, P. K. Nakane y G. B. Pierce en *Amer. J. Pathol*, 76: 549; 1974.

REMOVAL OF BASEMENT MEMBRANE IN THE INVOLUTING BREAST. A. Martínez-Hernández, L. M. Fink, y G. B. Pierce en *Lab. Invest.*, 34: 455; 1976.

GLUTAMINE SYNTHETASE: GLIAL LOCALIZATION IN BRAIN. A. Martínez-Hernández, K. P. Bell y M. D. Noremberg en *Science*, 195: 1356.

JUEGOS MATEMATICOS

TIME. C. D. Broad en *Encyclopaedia of Religion and Ethics*, dirigido por James Hastings. Charles Scribner's Sons, 1922.

TRAVELERS IN TIME: STRANGE TALES OF MAN'S JOURNEYINGS INTO THE PAST AND THE FUTURE. Dirigido por Philip Van Doren Stern. Doubleday & Co., Inc., 1947.

SYMBOLISM AND BELIEF. Edwyn Bevan. Kennikat Press, Inc., 1968.

TALLER Y LABORATORIO

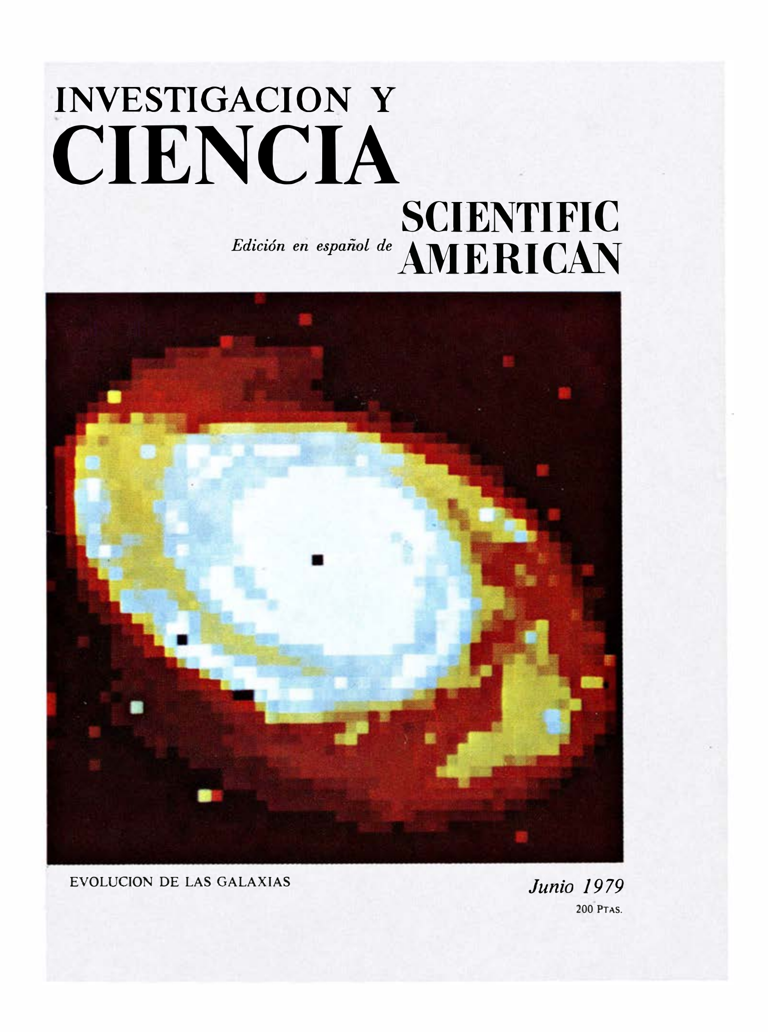
BOOMERANGS: MAKING AND THROWING THEM. Herb A. Smith. Gemstar Publications. Arun Sports, Littlehampton. Sussex. 1975.

MANY HAPPY RETURNS: THE ART AND SPORT OF BOOMERANGING. Benjamin Ruhe. The Viking Press, 1977.

ERRATUM

En la página 103 del número 30, marzo último, donde se lee ESTADO 1 debe leerse ESTADO 3, y a la inversa.

Seguiremos explorando los campos del conocimiento



TERATOMAS Y QUIMERAS, por Karl Illmensee y Leroy C. Stevens
Los teratomas son tumores formados por un conjunto de células y tejidos diferenciados. Sus células descendientes malignas pueden diferenciarse normalmente en el embrión del ratón y dar lugar a quimeras, o mosaicos genéticos.

EVOLUCION DE LOS TRASPLANTES DE CORNEA, por Joaquín Barraquer
La cirugía de la córnea ha progresado notablemente en el curso de los últimos 25 años. Pero deben investigarse nuevas técnicas y nuevas terapéuticas.

LA FISICA DEL KARATE, por Michael S. Feld, Ronard E. McNair y Stephen R. Wilk
El examen concienzudo de la forma con la que un experto en karate puede romper bloques de madera y de hormigón con su simple mano pone de manifiesto la notable capacidad del cuerpo humano para ejercer sin ayuda una fuerza física.

LA EVOLUCION DE LAS GALAXIAS DISCALES, por Stephen E. Strom y Karen M. Strom
Una galaxia espiral puede evolucionar a un disco suave sin brazos espirales. El que esto ocurra depende de su entorno: la galaxia más propensa a evolucionar a disco suave pertenece a un cúmulo rico.

ARMAS DE HACES DE PARTICULAS, por John Parmentola y Kosta Tsipis
Hay quien piensa en disparar misiles de acelerador. Aun cuando pudieran fabricarse tales ingenios, lo que es dudoso, no serían eficaces porque serían vulnerables a las contrarréplicas.

UNA ENZIMA DE LA GLANDULA PINEAL QUE CONTROLA EL TIEMPO, por Sue Binkley
La variación en la actividad de la enzima N-acetiltransferasa de la glándula pineal sirve de fundamento en algunos animales para un mecanismo natural que dicta el ritmo vital.

LA OSTEOMANCIA CHINA, por Hung-hsiang Chou
Los aristócratas de la dinastía Shang de hace 3000 años gustaban de solicitar el consejo de sus antepasados. Las preguntas, grabadas sobre hueso o sobre concha de tortuga, constituyen las claves para los orígenes de la civilización china.

LA ARCILLA, por Georges Millot
Tiene innumerables usos debido, sobre todo, a que está construida por finas partículas que son un reflejo de la naturaleza de sus minerales.

